DOCUMENT RESUME

ED 173 199 SO 011 751

AUTHOR Fox, Karen; And Others
TITLE An Analysis of the Relationship Between NCSS Curriculum Guidelines and Social Studies Achievement Tests.
INSTITUTION National Council for the Social Studies, Washington, D.C.
PUB DATE Nov 78
NOTE 86p.; Charts may not reproduce in paper copy from EDRS due to small print type of original document

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS *Academic Achievement; Achievement Tests; Educational Assessment; Educational Objectives; Educational Research; Elementary Secondary Education; *Guidelines; Learning; *Relationship; *Social Studies; Test Construction; *Test Results; Test Validity

ABSTRACT

The report evaluates the accuracy with which achievement tests measure social studies learning. Research is based on test items derived from the Social Studies Curriculum Guidelines (1971) and the Multiethnic Guidelines (1976), both prepared by National Council for the Social Studies (NCSS). The major objective of the project was to help test developers, social studies supervisors, and other educators improve test development and selection techniques. The document is presented in eight sections. Section I describes the 1971 and 1976 NCSS guidelines. Section II explores activities of the NCSS Committee on Testing. Section III rates the NCSS guidelines for their intent and potential for state, local, and national educational assessment. Section IV explains how the testing committee developed appropriate guideline statements and criteria for large-scale assessment. Section V identifies general implications of the guidelines for large-scale assessment. Sections VI and VII discuss developing the test review process and reporting test results. The final section offers conclusions and recommendations including that NCSS can help overcome some mis-matches between specific guidelines and published tests by sharing mutual concerns with publishers, by collaborating with publishers on appropriate roles for test makers, test users, and curriculum developers, and by encouraging more enlightened use of evaluation processes and data. (DB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED173199

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Karen F. A. Fox

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

AN ANALYSIS OF THE RELATIONSHIP BETWEEN NCSS CURRICULUM GUIDELINES AND SOCIAL STUDIES ACHIEVEMENT TESTS

Presented

to

Anna Ochoa, President, National Council for the Social Studies

and the

Board of Directors, National Council for the Social Studies

BEST COPY AVAILABLE

By the Committee on Testing
Karen Fox, Northwestern University
Paul Williams, Virginia State Department of Education
Jerry Moore, University of Virginia

With the Assistance of:
John Napier, University of Georgia
Mary Friend Shepard, Indiana University
Lynn Winters, Lawndale (CA) High School

SO 011 751

TABLE OF CONTENTS

		Page
Introduction		ii
Section I	Preliminary Statement on the NCSS Guidelines (1971 and 1976)	1
Section II	Activities of the Committee on Testing	3
Section III	Analysis of NCSS Curriculum Guidelines (1971 and 1976)	6
Section IV	Development of Criteria for Test Analysis	8
Section V	General Implications of the Guidelines for Large-Scale Assessment	37
Section VI	Development of the Test Review Process	44
Section VII	Reporting the Results	48
Section VIII	Conclusions and Recommendations	61
Appendix I	Vitae of the Committee on Testing	
Appendix II	Guidelines grouped by appropriate assessment processes	
Appendix III	Summary description of Information Produced by the States	

INTRODUCTION

This report addresses a fundamental question in the evaluation of social studies learning: To what extent are widely-used social studies achievement tests measuring the important outcomes of social studies instruction? Starting from the Social Studies Curriculum Guidelines (1971) and the Multiethnic Guidelines (1976), the NCSS Committee on Testing derived a list of social studies outcomes which could be measured with paper and pencil tests. Selected achievement tests were then evaluated for their congruence with these outcomes, including the breadth of content and cognitive skills represented.

The results of the analyses are presented in this report. There is, however, no "approved list" or other rankings of tests. Schools will continue to select specific tests to match their instructional goals and testing purposes. Our aim is to provide test developers, social studies supervisors, and other educators with information to make improved decisions in test development and selection.

The Committee on Testing was established by 1978 NCSS President Anna Ochoa, who saw the need for the social studies profession to communicate with those who develop and use tests. She was concerned to assure that significant social studies outcomes are included in testing programs. She had the guiding role of setting the Committee's task and encouraging us in our work.

Members of the Council of State Social Studies Supervisors participated in our survey of state testing programs to determine what tests were being used and what state curriculum guidelines were used to guide testing programs.

John Napier, Mary Friend Shepard, and Lynn Winters derived the learning outcomes from NCSS statements. Walter McPhie (University of Utah) and Allan Brandhorst (University of South Carolina) participated in the pilot stage of this task.

Paul Williams and Jerry Moore jointly carried out the analyses of tests and prepared this report, which embodies their broad knowledge of social studies education, their expertise in testing, and many weeks of their time.

This report was reviewed and approved by the Board of Directors of the National Council for the Social Studies in November 1978. The Committee on Testing hopes the report will not contribute to discussions of testing in the social studies.

Karen Fox, Coordinator
NCSS Ad Hoc Committee on Testing

TABLES

		Page
Table 1	NCSS Curriculum Guideline (1971): Appropriateness for Large-Scale Assessment	9-11
Table 2	NCSS Curriculum Guidelines (1971) By Level of Assessment and Type of Local Assessment	12
Table 3	NCSS Curriculum Guidelines (1976): Appropriateness for Large-Scale Assessment	13-17
Table 4	NCSS Curriculum Guidelines (1976) By Level of Assessment and Type of Local Assessment	18-20
Table 5	Content Relationships Between NCSS Guidelines, 1971 and 1976	21
Table 6	Distribution of Test Items on Norm-Referenced Social Studies Tests by Content Areas	51
Table 7	Distribution of Test Items on Norm-Referenced Social Studies Tests by Issue Orientation	52
Table 8	Distribution of Test Items on Norm-Referenced Social Studies Tests by Ethnic Content and Source of Ethnicity	53
Table 9	Distribution of Test items on Norm-Referenced Social Studies Tests by Recall vs. Higher Cognitive Skills and Pictorial/Reference Skills	54
Table 10	Distribution of Test Items on Norm-Referenced Social Studies Tests by Information Processing Skills	55
Table 11	Distribution of Test Items on State Social Studies Tests by Content Areas	61
Table 12	Distribution of Test Items on State Social Studies Tests by Issue Orientation	62
Table 13	Distribution of Test Items on State Social Studies Tests by Recall vs. Higher Cognitive Skills and Pictorial/Reference Skills	63
Table 14	Distribution of Test Items on State Social Studies Tests by Ethnic Content and Source of Ethnicity	64
Table 15	Distribution of Test Items on State Social Studies Tests by Information Processing Skills	65

CHARTS

		Page
Chart A	Factor 1: School Milieu	24
Chart B	Factor 2: Instructional Resources	25
Chart C	Factor 3: Instructional Development	26
Chart D	Factor 4: Student Participation	28
Chart E	Factor 5: Self-Worth	29
Chart F	Factor 6: Evaluation	30
Chart G	Factor 7: Scope and Sequence—Content	34–35
Chart H	Factor 8: Scope and Sequence—Skills	36

SECTION I

Preliminary Statement on The NCSS Guidelines (1971 and 1976)

The leadership of the NCSS appointed a task force to prepare guidelines for the social studies curriculum. In 1971 this NCSS Task Force on Curriculum Guidelines, in their report in Social Education, stated that the National Council was:

> "saying that after careful examination of recent developments in social studies education and in light of current developments in the larger society, the guidelines presented in this document need to be carefully considered in building a social studies program in the years immediately ahead." Social Education (December, 1971), p. 853.

The Task Force went on to assert that these curriculum guidelines should be helpful in the evaluation of various approaches to social studies education; in judging the soundness of social studies programs in schools and in determining the consistency of curriculum designs with recent thinking in social studies education.

The Task Force recognized that their work would be controversial to some while inadequate and ambiguous to others. Nevertheless, they developed guidelines that they felt would (should) be used by teachers, supervisors, and school officials to develop and evaluate social studies programs. Toward this end the Task Force on Curriculum Guidelines presented a three-part report including a rationale for social studies education, guidelines for social studies program development and an accompanying evaluation checklist.

In 1976, the NCSS Task Force of Ethnic Studies Curriculum Guidelines prepared a position statement for the NCSS entitled Curriculum Guidelines for Multiethnic Education. The preface to the document speaks clearly of the need for schools to embrace ethnic plurality within the school and

the society and to reduce the tensions produced by unexamined ethnic beliefs and attitudes. The position statement was divided into three secions: a rationale for ethnic pluralism, curriculum guidelines for multiethnic education and a checklist to evaluate school programs against the guidelines.

The 1971 and 1976 curriculum guidelines exhibit characteristics in common:

1) Both documents were produced to meet a perceived need resulting from events in the previous decade. The 1971 report was a response to the "cognitive" curriculum explosion in social studies education. The latter report was a response to the Civil Rights movement and rising consciousness about ethnic pluralism.

2) Both reports were attempts by the NCSS leadership to establish and assert the rights and responsibilities of the NCSS to influence and define the nature and practice of social studies education. Thus, both reports were political as well as professional in nature.

3) Finally, both reports suggested that the school and curricular elements in the guidelines could be used to evaluate school practices and programs.

At the time the 1971 Guidelines were published, the NCSS leadership acknowledged that many concerns could be raised about the Guidelines. At the same time it was recognized that these legitimate concerns would form the basis for further review, and potentially for changes in the document

itself. In fact, in the forward of the Guidelines it was stated that a three-member review committee would be established to monitor developments in the social studies for their potential impact upon the Guidelines. Consistent with this review process, the forward to the Guidelines stated that:

> During the next three years it is hoped and expected that the present document will be used, discussed, criticized, and thoroughly critiqued at the local, state, and national levels. Feedback from these deliberations can be used by the Review Committee to prepare a revised set of curriculum guidelines for the latter half of the 1970's. Social Education, p. 4 (1971)

The Committee on Testing perceives its present task of evaluating published social studies tests for their congruence to the 1971 and 1976 Guidelines as including, at least in part, an analysis and evaluation of the Guidelines themselves. We hope that this analysis is helpful to any future review process by which the Guidelines may be reviewed.

SECTION II

Activities of the Committee on Testing

On June 21, 1977 the NCSS Advisory Committee on Research was requested to undertake a project concerning testing in the social studies. NCSS President Anna Ochoa requested that Karen Fox, Research Committee Chairperson, assume responsibility for the completion of the project. Anna Ochoa created the NCSS Task Force on Testing (later renamed the NCSS Ad Hoc Committee on Testing) composed of Karen Fox, Paul Williams, and Jerry Moore. The initial charge given to the Committee on Testing (COT) included two primary tasks:

Task 1. The derivation of classes of learning outcomes based upon the 1971 and 1976 Guidelines, and

Task 2. The acquisition of a representative sample of commercially produced social studies tests to be analyzed for congruency with the outcomes derived above.

Walt McPhie, Linda Rickman, and Allan Brandhorse were requested to aid the COT in the completion of Task 1. Each person was directed to derive classes of learning outcomes based upon the Guidelines. They submitted their materials to the COT by the end of September, 1977.

After examining these materials it became evident that a more in-depth analysis of the Guidelines would be necessary before the COT could make substantial progress toward the completion of Task 2. The COT recognized that the Guidelines addressed highly variant issues in social studies education, consequently, each specific guideline statement was necessarily more or less amenable to large-scale assessment. These issues were discussed with NCSS President Anna Ochoa at the annual meeting in Cincinnati. As a result of this discussion, the first task was revised and expanded in order to allow for the creation of a metric by which an analysis of the published tests could be accomplished. The COT sought the help of selected social studies professionls to analyze the 1971 and 1976 Guidelines and to suggest the type of measurement and/or evaluation processes which would be consistent with each specific guideline. The three professionals were Lynn Winters, John Napier, and Mary Friend Shepard (vitae can be found in Appendix I). The work of these professionals was invaluable to the COT. First, they operated independently on what was a lengthy and difficult task. Secondly, as a group, they represented both secondary school and university

interests in social studies education. Third, their backgrounds complemented each other in breadth of interests and professional expertise. Finally, the group provided the COT with geographic representation.

A letter was sent to the group requesting that they provide the COT with the following information:

1. A narrative description of what tests should "look like" if they did in fact embody the Guidelines,

2. Item characteristics needed to satisfy the Guidelines,

3. Alternative assessment strategies, as long as they are amenable to large-scale assessment,

4. Content/processes which would be desireable on a test if it reflected the Guidelines, and

5. Possible examples of test items.

Napier, Shepard and Winters completed their task by March, 1978 and submitted their materials to the COT.

While work on Task 1 was proceeding, preliminary work was being accomplished for Task 2. The original plan was to collect only commercially published social studies tests and match them to the Guidelines. It became apparent, however, that it would be advantageous to examine what was being done in social studies testing at the state level. The reasons for this increased scope was the need to know the extent to which:

1. The states were using the NCSS Guidelines in establishing curriculum goals in social studies,

2. State evaluation processes in social studies met these goals, and

3. The extent to which the state tests contributed insights into social studies evaluation processes in general.

Thus, an initial letter was sent to CS 4 members (7/77) inquiring about the nature of their curricular and testing efforts in the social studies. Follow-up letters have been sent to gather information from as many states as possible. The results of this data collection are included in this report.

All major test publishers were requested to submit copies of social studies tests for review. These tests were reviewed and analyzed. The analysis of these tests comprises the final portion of this report.

SECTION III

Analysis of NCSS Curriculum Guidelines (1971 and 1976)

An initial analysis of the two NCSS curriculum documents revealed that specific guideline statements were unequal in importance and uneven in application for testing purposes. There were several reasons for this conclusion. First, the documents were addressed to needed changes in primary and secondary social studies programs. Therefore, the guidelines represented concerns broadly ranging from student learnings and behaviors to specifications for support systems essential to social studies programs. Secondly, the contents of the guidelines were prepared by a committee, reviewed by a committee and edited by another committee. Consequently, the specificity of any one guideline was highly variable across all guidelines. Finally, these curriculum documents were the first attempts to develop a curriculum philosophy for a professional group (NCSS) known for its

pluralism, its belief in conflict and debate and its long-standing search for the meaning of social studies education. It was not surprising to find both specific and ambiguous statements in the documents.

it was essential that the Committee on Testing (COT) establish quickly the extent to which each curriculum guideline statement was appropriate for large-scale assessment. The COT proceeded on this task as follows:

1. The reports of John Napier, Mary Friend Shepard and Lynn Winters were recorded on a chart that included each sepcific guideline for each the 1971 and 1976 documents (the ratings are reported in Tables 1 and 3).

2. The Committee on Testing discussed each guideline considering the interpretations and comments of Napier, Shepard and Winters.

3. Finally, the Committee on Testing also rated each guideline as appropriate or inappropriate for large-scale assessment along with comments on the type of assessment required or available (See Tables 1 and 3).

The composite ratings reported in Tables 1 and 3 produced many suggestions about the appropriate level of assessment--national, state and/or local agency--and the portential types of assessment for those guidelines deemed inappropriate for large-scale assessment. These data are reported in Tables 2 and 4.

Further, the ratings of the 1971 and 1976 NCSS Curriculum Guidelines allowed the Committee on Testing to share and examine personal feelings about the nature and uses of assessment devices. The reports and comments of Napier, Shepard and Winters provided some external controls on the

subjectivity of committee discussions.

The committee deliberations required that each guideline be interpreted for its intent and purpose and for its potential for large-scale and local assessment. Tables 1 and 3 represent a distillation of the committee's deliberations. A by-product of these discussions was a re-grouping of the individual guidelines (1971; 1976) under the following headings:

National Assessment
State Assessment
Local Assessment - Unobtrusive
Local Assessment - Attitudinal
Local Assessment - Checklists

These data are reported in Tables 2 and 4.

SECTION IV

Development of Criteria For Test Analysis

After each of the guidelines had been rated for its appropriateness for large-scale assessment, the Committee on Testing acted to determine the redundancies between the 1971 NCSS Curriculum Guidelines and the 1976 NCSS Multiethnic Guidelines. This task was important to:

1) Develop the simplest list of guideline statements that were appropriate for large-scale assessment.

2) Develop a set of criteria that formed the basis for a metric by which published social studies tests could be judged.

The 1971 and 1976 NCSS Curriculum Guidelines were juxtaposed on a bivariate grid (see Table 5) permitting continuous and cross-reference

TABLE 1
NCSS CURRICULUM GUIDELINES (1971)
APPROPRIATENESS FOR LARGE-SCALE ASSESSMENT
(No=Inappropriate) (Yes=Appropriate)

Guide-line Ref.	Raters				Comments
	Napier	Shepard	Smith	COT	
1.1	no	no	no	no	This category provides for a checklist of curriculum and curriculum development but not assessment of student outcomes.
1.2	no	no	no	no	
1.3	no	no	no	no	
1.4	no	no	no	no	
2.1	yes	yes	yes	yes	
2.2	yes	yes	yes	yes	
2.3	yes	yes	yes	yes	
2.4	yes	yes	yes	yes	
2.5	yes	yes	no		Cues on test items could show ethnicity
2.6	no	no	no	no	These are localized values.
2.7	no	no	no	no	Unobtrusive measures.
3.1	yes	yes	yes	yes	Tests need flexibility—build item banks. Could report results by cognitive level. NCSS needs to define clearly.
3.2	yes	yes	yes	yes	
3.3	yes	yes	yes	yes	
3.4	no	no	yes	yes	Defines the scope and breadth of the social studies. Related to structure of traditional disciplines.
3.5	no	no	yes	yes	
3.6	no	no	yes	yes	
3.7	yes	yes	yes	yes	
3.8	no	yes	yes	yes	
4.1	no	no	no	no	Allows for comparison of curriculum goals to instructional plans.
4.2	no	no	no	no	
4.3	no	no	no	no	
4.4	no	no	no	no	

TABLE 1 (continued)
NCSS CURRICULUM GUIDELINES (1971)
APPROPRIATENESS FOR LARGE-SCALE ASSESSMENT
(No=Inappropriate) (Yes=Appropriate)

Guide-line Ref.	Raters				Comments
	Napier	Shepard	Smith	COT	
4.5	no	no	no	no	
4.6	no	no	no	no	
5.1	no	no	no	no	
5.2	yes	yes	yes	yes	See 3.2 Methods of Inquiry.
5.3	yes	yes	no	yes	Partially under 3.2.
5.4	no	no	no	no	
5.5	no	no	no	no	No student outcomes.
5.6	no	no	no	no	
5.7	no	no	yes	no	Self-concept (nominally).
6.1	no	no	no	no	
6.2	no	no	no	no	
6.3	no	no	no	no	Instructional Resources.
6.4	no	no	no	no	
6.5	no	no	no	no	
6.6	no	no	no	no	
7.1	no	no	yes	yes	See Michigan materials.
7.2	no	no	yes	yes	
7.3	no	no	no	no	
7.4	no	no	no	no	
8.1	—	—	—	—	General discussion of the use of tests, results and test users.
8.2	—	—	—	—	
8.3	—	—	—	—	

TABLE 1 (continued)
NCSS CURRICULUM GUIDELINES (1971)
APPROPRIATENESS FOR LARGE-SCALE ASSESSMENT
(No=Inappropriate) (Yes=Appropriate)

Guide-line Ref.	Raters				Comments
	Napier	Shepard	Smith	COT	
8.4	—	—	—	—	
8.5	—	—	—	—	
8.6	—	—	—	—	
8.7	—	—	—	—	
8.8	—	—	—	—	
9.1	no	no	no	no	
9.2	no	no	no	no	
9.3	no	no	no	no	
9.4	no	no	no	no	
9.5	no	no	no	no	
9.6	no	no	no	no	
9.7	no	no	no	no	

TABLE 2
NCSS CURRICULUM GUIDELINES (1971)
BY LEVEL OF ASSESSMENT AND TYPE OF
LOCAL ASSESSMENT

Guideline Reference	Large-Scale Assessment (National)	Large-Scale Assessment (State)	Appropriate Types of Local Assess.			Notes
			Unobtrusive Measures	Attitude Surveys	Checklists	
1.1-1.4			x		x	Studend participation
2.1-2.4	x	x				
2.5				x		Picture cues show ethnicity
2.6		x	x			
2.7			x		x	
3.1-3.8	x	x				Defines scope of SS--need to report skill results accurately
4.1-4.6			x		x	Compare instruction to curriculum
5.1			x		x	
5.2-5.3	x	x				Subsumed under 3.2
5.4-5.5			x			
5.6-5.7			x	x		Self-concept
6.1-6.6			x		x	Instructional plans
7.1-7.2	x	x				Highest cognitive evaluation
7.3			x			

TABLE 3
NCSS CURRICULUM GUIDELINES (1976)
APPROPRIATENESS FOR LARGE-SCALE ASSESSMENT
(No=Inappropriate) (Yes=Appropriate)

Guide-line Ref.	Raters				Comments
	Napier	Shepard	Winters	COT	
1.0	—	no	no	no	Concern for test writer values
1.1	no	no	no	no	
1.2	no	no	no	no	
1.3	no	no	no	no	
1.4	no	no	no	no	
1.5	no	no	no	no	
2.0	—	no	no	no	Localized issues
2.1	no	no	no	no	
2.2	no	no	no	no	
2.3	no	no	no	no	
2.4	no	no	no	no	
2.5	no	no	yes	no	
3.0	—	no	no	no	
3.1	no	no	no	no	
4.0	—	no	no	no	In-service development issues.
4.1	no	no	no	no	
4.2	no	no	no	no	
4.3	no	no	no	no	
4.4	no	no	no	no	
4.5	no	no	no	no	
4.6	no	no	no	no	
5.0	—	no	no	no	NCSS must define basic terms.

TABLE 3 (continued)
NCSS CURRICULUM GUIDELINES (1976)
APPROPRIATENESS FOR LARGE-SCALE ASSESSMENT
(No=Inappropriate) (Yes=Appropriate)

Guide-line Ref.	Raters				Comments
	Napier	Shepard	Winters	COT	
5.1	yes	no	yes	no	
5.2	no	no	no	no	
6.0	—	no	no	no	
6.1	yes	yes	yes	no	Discuss psych tests of self-concept
6.2	yes	yes	yes	no	
6.3	yes	yes	yes	yes	
6.4	yes	yes	yes	yes	
7.0	—	no	no	no	
7.1	yes	yes	yes	yes	Concerns test content—i.e. racism.
7.2	yes	yes	yes	yes	
7.3	yes	yes	yes	yes	
7.4	yes	yes	yes	yes	
7.5	yes	yes	no	yea	Concern for stimulus content
7.6	yes	yes	no	yes	
7.7	yes	yes	no	yes	
8.0	—	no	no	no	
8.1	yes	yes	yes	yes	Issues of inquiry—item content
8.2	yes	yes	yes	yes	
9.0	—	no	no	no	See 3.0 NCSS curriculum guidelines
9.1	no	no	no	no	
9.2	no	no	no	no	
10.0	—	no	no	no	See 2.0 NCSS curr. guidelines

TABLE 3 (continued)
NCSS CURRICULUM GUIDELINES (1976)
APPROPRIATENESS FOR LARGE-SCALE ASSESSMENT
(No=Inappropriate) (Yes=Appropriate)

Guide-line Ref.	Raters				Comments
	Napier	Shepard	Winters	COT	
10.1	yes	yes	yes	yes	Could be done by attitude surveys
10.2	yes	yes	yes	yes	
10.3	no	no	no	no	
10.4	yes	yes	no	no	
11.0	—	no	yes	yes	Really explicates 5.0 NCSS curr. guidelines
11.1	yes	yes	yes	yes	
11.2	yes	yes	yes	yes	
11.3	yes	yes	yes	yes	
11.4	yes	yes	yes	yes	
11.5	yes	yes	yes	yes	
11.6	yes	yes	yes	yes	
11.7	no	no	no	no	
11.8	yes	yes	yes	yes	
12.0	—	no	no	no	
12.1	yes	no	no	no	General concensus
12.2	no	no	no	no	
13.0	—	no	no	no	
13.1	yes	yes	no	no	
13.4	no	no	no	no	
13.5	no	no	no	no	
13.6	no	no	no	no	
13.7	no	no	no	no	

TABLE 3 (continued)
NCSS CURRICULUM GUIDELINES (1976)
APPROPRIATENESS FOR LARGE-SCALE ASSESSMENT
(No=Inappropriate) (Yes=Appropriate)

Guide-line Ref.	Raters				Comments
	Napier	Shepard	Winters	COT	
14.0	—	no	no	no	Knowledge of these groups included
14.1	yes	yes	no	no	
14.2	yes	yes	no	no	
14.3	yes	yes	no	no	
15.0	—	no	no	no	
15.1	no	no	no	no	
15.2	no	no	no	no	
16.1	yes	yes	no	no	Attitudinal or judgmental
16.2	yes	yes	no	no	
17.0	—	no	yes	yes	
17.1	no	no	yes	yes	
17.2	yes	yes	yes	yes	
17.3	yes	yes	yes	yes	
18.0	—	no	no	no	See 2.0 NCSS curr. guidelines
18.1	yes	yes	yes	yes	
18.2	yes	yes	yes	yes	
19.0	—	no	no	no	Local checklist/or cognitive assessment
19.1	no	no	no	no	
19.2	no	no	no	no	
19.3	yes	yes	yes	yes	Content for 2.0 NCSS
19.4	no	no	yes	yes	Curriculum guidelines
19.5	no	no	no	no	

TABLE 3 (continued)
NCSS CURRICULUM GUIDELINES (1976)
APPROPRIATENESS FOR LARGE-SCALE ASSESSMENT
(No=Inappropriate) (Yes=Appropriate)

Guide-line Ref.	Raters				Comments
	Napier	Shepard	Winters	COT	
19.6	yes*	no	no	no	20.0 integration vs salad bowl
20.0	__	no	no	no	
20.1	yes*	yes	yes	yes	
20.2	yes*	yes	no	no	Calls for a local report
20.3	yes*	yes	yes	yes	
20.4	no	no	no	no	Worthwhile knowledge?
21.0	__	no	no	no	Various types of checklists
21.2	no	no	no	no	
21.3	no	no	no	no	
22.0	__	no	no	no	
22.1	no	no	no	no	
22.2	no	no	no	no	Various types of assessment conducted in local schools
23.0	__	no	no	no	
23.1	no	no	no	no	
23.2	no	no	no	no	
23.3	no	no	no	no	
23.4	no	no	no	no	
23.5	no	no	no	no	

*Identified as measurable student outcomes but deemed outside the social studies domain.

TABLE 4
NCSS CURRICULUM GUIDELINES (1976)
BY LEVEL OF ASSESSMENT AND TYPE OF
LOCAL ASSESSMENT

Guideline Reference	Large-Scale Assessment (National)	Large-Scale Assessment (State)	Appropriate Types of Local Assess.			Notes
			Unobtrusive Measures	Attitude Surveys	Checklists	
1.0						
1.1-1.2						
1.3-1.5			x		x	
2.0						
2.1-2.5		x	x	x	x	
3.0-3.1			x	x	x	
4.0-4.6			x	x	x	
5.0-5.2	—	—	—	—	—	Suggestions for test construction
6.0-6.4		x	x	x		Psychometric tests on self-concept/ state efforts
7.0-7.4	x	x				NCSS Curr. Guideline 2.0 examine ethnic content
7.5-7.7	x	x	x			Stimulus of test items
8.0-8.2	x	x	x			Skills section 3.0 NCSS curr. guidelines
9.0	x	x	x			NCSS curr. guildelines 3.0

TABLE 4 (continued)
NCSS CURRICULUM GUIDELINES (1976)
BY LEVEL OF ASSESSMENT AND TYPE OF
LOCAL ASSESSMENT

Guideline Reference	Large-Scale Assessment (National)	Large-Scale Assessment (State)	Appropriate Types of Local Assess.			Notes
			Unobtrusive Measures	Attitude Surveys	Checklists	
9.1-9.2						
10.0-10.1	x	x				Explication of NCSS curriculum guideline 2.0
10.2-10.4			x	x	x	
11.0-11.6	x	x	x			Explication of NCSS curriculum guideline 5.0
11.7-11.8		x	x	x	x	
12.0-12.1	x	x				Item produce reflects ethnic knowledge
12.2			x		x	
13.0-13.3	x	x				Refer NCSS curr. guideline 3.0
13.4-13.7			x	x	x	
13.6	x**					Only reference to a hierarchy of concepts
14.0-14.3	x**		x	x	x	Knowledge of ethnic groups
15.0-15.2			x		x	Instructional management

TABLE 4 (continued)
NCSS CURRICULUM GUIDELINES (1976)
BY LEVEL OF ASSESSMENT AND TYPE OF
LOCAL ASSESSMENT

Guideline Reference	Large-Scale Assessment (National)	Large-Scale Assessment (State)	Appropriate Types of Local Assess.			Notes
			Unobtrusive Measures	Attitude Surveys	Checklists	
16.0-16.2	x	x				Skills found under NCSS curriculum guideline 3.0
17.0-17.1			x		x	Curriculum analysis
17.2-17.3	x	x	x	x		Higher cognitive level--NCSS curriculum guideline 3.0
18.0-18.2	x	x				NCSS curriculum guideline 2.0
19.0-19.2			x		x	Could refer to the use of primary sources
19.3-19.4	x	x				NCSS curriculum guidelines 2.0
19.5-19.6			x	x	x	
20.0-20.1	x	x				NCSS curriculum guideline 2.0 language transforms
20.2-20.3			x	x	x	Question dialect objective here
20.4			x	x	x	
21.0-21.3			x	x	x	
22.0-22.2		x	x			Implications for the use of tests
23.0-23.5		x			x	Sampling of teacher built tests

TABLE 5
FACTOR ANALYSIS
NCSS GUIDELINES 1971 and 1976

1976 Guidelines \ 1971 Guidelines	1.0 (1.1-.4)	2.0 (2.0-.2)	(2.3)	(2.4)	(2.5)	(2.6)	(2.7)	3.0 (3.1)	(3.2-.3)	(3.4-.6)	4.0 (4.1-.6)	5.0 (5.1)	(5.2-.3)	(5.4-.7)	6.0 (6.1-.6)	7.0 (7.1-.2)	(7.3-.4)	8.0 (8.1-.8)	9.0 (1.1-.7)	FACTOR NAMES
1.0						X									X					Instr. Resources
(1.1)		X		X				X		X										S/S-Content
(1.2-.5)						X									X					Instr. Resources
2.0 (2.1-.5)																			X	School Milieu
3.0 (3.1)																			X	School Milieu
4.0 (4.1-.6)																			X	School Milieu
5.0 (5.1-.2)												X	X		X			X		Instr. Dev.
6.0 (6.1-.4)	1																			Self-Worth[1]
7.0 (7.1-.7)		X		X				X		X										S/S-Content
8.0 (8.1-.2)													X			X				S/S-Skills
9.0 (9.1-.2)		X		X				X		X										S/S-Content
10.0 (10.1-.4)		X		X				X		X										S/S-Content
11.0 (11.1-.6)													X			X				S/S-Skills
(11.7)	X				X		X													St. Participation
(11.3)																				Self-Worth[1]
12.0 (12.1-.2)																				Self-Worth[1]
13.0 (13.1-.2)		X		X				X		X										S/S-Content
(13.3-.5)						X									X					Instr. Resources
(13.6)													X			X				S/S-Skills
14.0 (14.1-.3)		X		X				X		X										S/S-Content
15.0 (15.1-.2)		X		X				X		X										S/S-Content
16.0 (16.1-.2)													X			X				S/S-Skills
17.0 (17.1-.2)		X		X				X		X										S/S-Content
(17.3)													X			X				S/S-Skills
18.0 (18.1-.2)		X		X				X		X										S/S-Content
19.0 (19.1-.2)																			X	School Milieu
(19.3-.6)		X		X				X		X										S/S-Content
20.0 (20.1-.4)		X		X				X		X										S/S-Content
21.0 (21.1-.3)						X									X					Instr. Resources
22.0 (22.1-.2)																		X		Evaluation
23.0 (23.1-.5)																		X		Evaluation

FACTOR NAMES
St. Part.
S/S-Content
S/S-Skills
S/S-Content
St. Part.
Instr. Res.
St. Part.
S/S-Content
S/S-Skills
S/S-Content
Instr. Dev.
Instr. Dev.
S/S-Skills
Instr. Dev.
Instr. Res.
S/S-Skills
Instr. Dev.
Evaluation
School Milieu

[1]NOTE: Self-worth relates closely with student participation under 1.0 NCSS Guidelines (1971).

reporting of guideline entries. (NOTE: the Guidelines, where similar, were grouped in the table to make for simpler charting). This analysis allowed the committee to identify and name selected factors present in both the 1971 and 1976 documents. The major factors identified were:

1) School Milieu
2) Instructional Resources
3) Instructional Development
4) Student Participation
5) Self-worth
6) Evaluation
7) Scope and Sequence-Content
8) Scope and Sequence-Skills

The broadly differentiated qualities of these factors clearly demonstrated that many segments of the 1971 and 1976 Curriculum Guidelines were never intended for large-scale assessment. The Committee on Testing had to define "curriculum" if we were to successfully assess the relationship between published social studies tests and the goals of social studies education imbedded in the Curriculum Guidelines. By defining "curriculum" the Committee on Testing could be successful in rejecting/retaining the factors for further study.

The Committee on Testing defined "curriculum" as "a set of intended learning outcomes." This definition was deemed particularly appropriate for the examination of large-scale assessment instruments. The Committee on Testing proceeded to discuss the inclusion or exclusion of each of the factors identified in the Curriculum Guidelines. Summary discussions of the excluded factors follow:

FACTOR 1: School Milieu

The specific guidelines isolated on this factor are reported on Chart A. Essentially, the composition of these guideline statements are: (1) staff development of those school staffs essential to the social studies

programs; (2) school policies supportive of social studies programs; (3) school support for aesthetic experiences and (4) support (time and money) for social studies staff. While Factor 1 contributed positively to social studies program development, its composition provided no insights into the learning outcomes of a social studies curriculum. Therefore, the school milieu factor was rejected in the examination of large-scale assessment of social studies.

FACTOR 2: Instructional Resources

The NCSS Curriculum and Multiethnic Guidelines associated with this factor are reported on Chart B. These guideline statements may be characterized as (1) physical resources available for social studies instruction; (2) media resources appropriate for multiethnic, social studies instruction and (3) the use of the community in the school. While all of these conditions contribute to the quality of social studies instruction, they do not form criteria that can be used to analyze large-scale assessment instruments in social studies.

FACTOR 3: Instructional Development

The guideline statements classified under this factor are reported in Chart C. These NCSS Curriculum Guidelines (1971, 1976) were characterized as: (1) directions for lesson planning; (2) attitudes toward learning; (3) differences in learning styles and (4) interdisciplinary and multidisciplinary dimensions of instruction. The Committee on Testing judged that these guidelines promoted the articulation of a social studies curriculum, but could not be used to directly analyze large-scale assessments in social studies.

FACTOR 4: Student Participation

The NCSS Guidelines assigned to this factor appear on Chart D.

CHART A

Factor 1: School Milieu

Guidelines 1976	Guidelines 1971
2.0 Do school policies and procedures foster positive interactions among the different ethnic group members of the school?	9.1 Does the school provide appropriate materials, time, and facilities for social studies education? R T C
2.1 Do school policies accommodate the behavioral patterns, learning styles, and orientations of those ethnic group members actually in the school?	9.2 Do teachers try out and adapt for their own students promising innovations? R T
2.2 Does the school provide a diversity of instruments and techniques in teaching and counseling students of different ethnic groups?	9.3 Are the basic purposes of social studies education as clearly related to the needs of the immediate community as to those of society at large? R C
2.3 Do school policies recognize the holidays and festivities of different ethnic groups?	9.4 Do teachers participate regularly in active social studies curriculum committees with both decision-making and advisory responsibilities? T
2.4 Do school policies avoid instructional and guidance practices based on stereotyped and ethnocentric perceptions?	9.5 Do teachers participate regularly in activities which foster their competence in social studies education? R T
2.5 Do school policies respect the dignity and worth of students as individuals *and* as members of ethnic groups?	9.6 Do teachers have social studies consultants available for help? R T
3.0 Are the school staffs (administrative, instructional, counseling, and supportive) multiethnic and multiracial?	9.7 Can teachers and schools rely upon a district-wide policy statement on academic freedom and professional responsibility? R T C
3.1 Has the school established and enforced policies for recruiting and maintaining multiethnic, multiracial staffs?	
4.0 Does the school have systematic, comprehensive, mandatory, and continuing multiethnic staff development programs?	
4.1 Are teachers, librarians, counselors, administrators, and the supportive staff included in the staff development programs?	
4.2 Do the staff development programs include a variety of experiences (such as lectures, field experiences, curriculum projects, etc.)?	
4.3 Do the staff development programs provide opportunities to gain knowledge and understanding about different ethnic groups?	
4.4 Do the staff development programs provide opportunities for participants to explore their attitudes and feelings about their own ethnicity and others'?	
4.5 Do the staff development programs examine the verbal and nonverbal patterns of interethnic group interactions?	
4.6 Do the staff development programs provide opportunities for learning how to create and select multiethnic instructional materials and how to incorporate ethnic content into curriculum materials?	
19.0 Does the school provide opportunities for students to participate in the aesthetic experiences of various ethnic groups?	
19.1 Are multiethnic literature and art used to promote empathy for people of different ethnic groups?	
19.2 Are multiethnic literature and art used to promote self-examination and self-understanding?	

CHART B

Factor 2: Instructional Resources

Guidelines 1976

1.0 Does ethnic pluralism permeate the total school environment?

1.2 Do instructional materials treat ethnic differences and groups honestly, realistically, and sensitively?

1.3 Do school libraries and resource centers have a variety of materials on the histories, experiences, and cultures of many different ethnic groups?

1.4 Do school assemblies, decorations, speakers, holidays, and heroes reflect ethnic group differences?

1.5 Are extracurricular activities multiracial and multiethnic?

13.3 Does the curriculum include the role of ethnicity in the local community as well as in the nation?

13.4 Does content related to ethnic groups extend beyond special units, courses, occasions, and holidays?

13.5 Are materials written by and about ethnic groups used in teaching fundamental skills?

21.0 Does the curriculum make maximum use of local community resources?

21.1 Are students carefully involved in the continuous study of the local community?

21.2 Are members of the local ethnic communities continually used as classroom resources?

21.3 Are field trips to the various local ethnic communities provided for students?

Guidelines 1971

2.6 Does the program build upon the realities of the immediate school community? R T C

6.1 Does the program have a wealth of appropriate instructional resources? R T O

6.2 Do printed materials accommodate a wide range of reading abilities and interests, learning activities, and sources? R O

6.3 Is a variety of media available for learning through many senses? R O

6.4 Do classrooms draw upon the contributions of many kinds of resource persons and organizations representing many points of view? T O C

6.5 Do activities use the school and community as a learning laboratory? T O C

6.6 Does the program have available many kinds of work space? R O

CHART C

Factor 3: Instructional Development

Guidelines 1976	Guidelines 1971
5.0 Does the curriculum reflect the ethnic learning styles of students within the school?	4.1 Are objectives carefully selected and formulated? R T
5.1 Is the curriculum designed to help students learn how to function effectively in different cultural environments and master more than one cognitive style?	4.2 Are knowledge, abilities, valuing, and social participation all represented in the objectives of the program? R T
5.2 Do the objectives, instructional strategies, and learning materials reflect the cultures and cognitive styles of the different ethnic groups within the school?	4.3 Are general statements of goals translated into specific objectives conceived in terms of behavior and content? R T
13.7 Is the study of ethnicity incorporated in instructional plans rather than being supplementary or additive?	4.4 Are classroom instruction and materials based upon clearly stated objectives? R T O
15.0 Are interdisciplinary and multidisciplinary approaches used in designing and implementing the multiethnic curriculum?	4.5 Does classroom instruction enable students to see their goals clearly in brief instructional sequences and lengthy units of study? T S
15.1 Are interdisciplinary and multidisciplinary perspectives used in the study of ethnic groups and related issues?	4.6 Are objectives reconsidered and revised periodically? R T
15.2 Are approaches used authentic and comprehensive explanations of ethnic issues, events, and problems?	5.1 Do students have a wide and rich range of learning activities appropriate to the objectives of their program?
	5.4 Do activities involve students in their communities? S T C
	5.5 Are learning activities sufficiently varied and flexible? T S
	5.6 Do students perceive their teachers as fellow inquirers? S
	5.7 Are activities carried on in a climate which supports students' self-respect and opens opportunities to all?
	7.3 Does the program enable students to relate their experiences in social studies to other areas of experience?
	7.4 Does the formal pattern of the program offer choice and flexibility? R S

Essentially, these guideline statements focus upon student involvement in (1) the instructional process and (2) community action. The 1971 NCSS Task Force specifically discussed the importance of social studies education beyond the classroom (Social Education, December, 1971, p. 859). While these guidelines are important to the social studies program, they do not contain intended student behaviors that could be included on large-scale assessments.

FACTOR 5: Self-Worth

The 1976 Multiethnic Guidelines contain several statements that emphasize the development of self. These statements are recorded on Chart E. It can be inferred that section 1.0 of the 1971 Guidelines (reported under student participation above) contribute heavily to this factor. The Committee on Testing noted that there are many non-content specific instruments available for the assessment of self-worth. However, these instruments are not appropriate for the direct assessment of social studies curriculum goals and outcomes.

FACTOR 6: Evaluation

Guidelines from both the 1971 and 1976 documents relating to evaluation processes and products appear in Chart F. These guideline statements provide thorough direction for local and large-scale assessment procedures, especially directions for objective referenced testing. Characteristics of this factor are discussed under 5.0 of this report.

Two factors were identified as appropriate for the consideration of large-scale assessment strategies. These factors were named (1) Scope and Sequence-Content and (2) Scope and Sequence-Skills. The Committee on Testing judged that these two factors were amenable to large-scale assessment techniques and reflected student outcomes consistent with its definition of

CHART D

Factor 4 : Student Participation

Guidelines 1976	Guidelines 1971
11.7 Does the curriculum provide opportunities for students to take action on social problems affecting ethnic groups?	1.1 Are students involved in the formulation of goals, the selection of activities, and the assessment of curriculum outcomes? S T 1.2 Do the school and its teachers make steady effort, through regularized channels and practices, to identify areas of concern to students? S T 1.3 Do students have choices within programs? S T 1.4 Do all students have ample opportunity for social studies education at all grade levels? R T 2.5 Does the program offer opportunities to meet and work with members of racial and ethnic groups other than their own? R T C 2.7 Is participation both in school and out considered a part of the program? P S C

CHART E

Factor 5: Self-Worth

Guidelines 1976	Guidelines 1971
6.0 Does the curriculum provide continuous opportunities for students to develop a better sense of self? 6.1 Does the curriculum help students strengthen their self-identities? 6.2 Is the curriculum designed to help students develop greater self-understanding? 6.3 Does the curriculum help students improve their self-concepts? 6.4 Does the curriculum help students better understand themselves in the light of their ethnic heritages? 11.8 Does the curriculum help students develop a sense of efficacy? 12.0 Does the curriculum help students develop skills necessary for effective interpersonal and inter-ethnic group interactions? 12.1 Does the curriculum help students understand ethnic reference points which influence communication? 12.2 Does the curriculum help students try out cross-ethnic experiences and reflect upon them?	NOTE: Some inference of self-worth is contained under guidelines 1.0, 1.1, 1.2, 1.3 and 1.4, which we classified as student participation.

CHART F

Factor 6: Evaluation

Guidelines 1976

22.0 Do the assessment procedures used with students reflect their ethnic cultures?

22.1 Do teachers use a variety of assessment procedures which reflect the ethnic diversity of the students?

22.2 Do teachers' day-to-day assessment techniques take into account the ethnic diversity of the students?

23.0 Does the school conduct ongoing, systematic evaluations of the goals, methods, and instructional materials used in teaching about ethnicity?

23.1 Do assessment procedures draw on many sources of evidence from many sorts of people?

23.2 Does the evaluation program examine school policies and procedures?

23.3 Does the evaluation program examine the everyday climate of the school?

23.4 Does the evaluation program examine the effectiveness of curricular programs, academic and non-academic?

23.5 Are the results of evaluation used to improve the school program?

Guidelines 1971

8.1 Is evaluation based primarily on the school's own statements of objectives? R

8.2 Does assessment include progress in knowledge, abilities, valuing, and participation? R

8.3 Does evaluation data come from many sources, inside and outside the classroom? R

8.4 Are evaluation procedures regular, comprehensive, and continuous? R

8.5 Are evaluation data used for planning curricular improvement? R T

8.6 Do evaluation data offer students help in the course of learning? T S

8.7 Are both students and teachers involved in the process of evaluation? S T

8.8 Is regular re-examination of basic curricular goals an integral part of the evaluation? R T

curriculum.

The Committee on Testing examined the Scope and Sequence-Content factor first--hoping that these guidelines would best describe how the 1971 and 1976 Guideline Committees defined "social studies education." (NOTE: the COT found Bulletin 51 NCSS to be very helpful.) A series of questions were directed toward this factor:

> What is the conceptual nature of social studies education?
>
> What are the content (discipline) areas to be included in the social studies curriculum?
>
> What issues (topics), if any, are constant in the social studies curriculum?
>
> What sequencing of content, if any, should be considered within the social studies curriculum?

The Scope and Sequence-Content factor is presented in Chart G. Initial observations suggest that the 1971 and 1976 Guidelines are highly redundant. Both sets of guidelines share common sources of curriculum breadth, orientation toward pervasive issues (problems) and concerns for ethnic diversity.

The Committee on Testing discussed each of the guidelines reported under the Scope and Sequence-Content factor in order to produce criteria for the examination of published social studies tests. The following statements are summaries of the content factor:

1) The social studies curriculum should draw content from history and the social sciences with some representation from related fields of law, psychology, etc.

2) The social studies curriculum should include enduring issues (problems) as reflected in

contemporary society.

3) Within the framework of the traditional disciplines and social issues, the social studies curriculum must include holistic views of ethnic groups--concepts, facts and values. Tokenism should be eliminated.

While most of the Scope and Sequence-Content guidelines were helpful in generating the above criteria, there were some statements that had limited utilities for the task. For example, Guidelines 2.1 (1971) states "Does the program focus on the social world as it actually is?"; (1971) states "Does the program include a careful selection of that knowledge of most worth?"; 10.0 (1976) states "Does the curriculum promote values, attitudes and behavior which support ethnic pluralism?". These statements represented value assumptions that lacked clarity. Who knows what the social world really is? How a valid concept differs from an invalid one? and which knowledge is of most worth?

The Scope and Sequence-Skills factor was discussed last by the Committee. The summary of the 1971 and 1976 Guidelines on this factor appears in Chart H. During the discussions the following questions were directed toward the skills factor:

> What operational differences exist between inquiry, reflective thinking, valuing, decision-making and critical thinking skills?
>
> Is there a hierarchy (or hierarchies) of social studies skills?
>
> What (if any) skills are essentially the domain of social studies education?

The Committee on Testing produced the following summary statements about the skills factor:

1) The social studies curriculum should provide for the development of procedural skills—reference, location, graph, visual, tabular and chronological skills.

2) The social studies curriculum should provide for the development of information processing skills—e.g. inference, compare/contrast, classification, hypothesis/generalization, interpretation skills, etc.

3) The social studies curriculum should provide for a hierarchy (hierarchies) of skill development—simple to complex.

4) Ethnic content and concepts should be used frequently as the substance for the practice and development of social studies skills.

While the 1971 and 1976 Guidelines were helpful in producing the above criteria, the statements' lack of specificity and cohesion did not provide sufficient direction. For example, methods of inquiry, decision-making and/or valuing are not defined adequately enough to allow differentiation between them. The Multiethnic Guidelines (1976) are somewhat superior in this regard—they specify skills such as discriminating between fact and opinion—but even these guidelines lack a coherent framework.

CHART G

Factor 7: Scope and Sequence—Content

Guidelines 1976

1.1 Is ethnic content incorporated into all aspects of the curriculum, preschool through grade twelve and beyond?

7.0 Does the curriculum help students to understand the wholeness of the experiences of ethnic groups?

7.1 Does the curriculum include the study of societal problems some ethnic group members experience, such as racism, prejudice, discrimination, and exploitation?

7.2 Does the curriculum include the study of historical experiences, cultural patterns, *and* social problems of different ethnic groups?

7.3 Does the curriculum include both positive and negative aspects of ethnic group experiences?

7.4 Does the curriculum present ethnics as active participants in society *and* as subjects of oppression and exploitation?

7.5 Does the curriculum examine the diversity within each ethnic group's experience?

7.6 Does the curriculum present ethnic group experience as dynamic and continuously changing?

7.7 Does the curriculum examine the experiences of ethnic group people instead of focusing exclusively on the "heroes"?

9.0 Does the curriculum explore and clarify ethnic alternatives and options within American society?

9.1 Does the teacher create a classroom atmosphere reflecting an acceptance of and respect for ethnic differences?

9.2 Does the teacher create a classroom atmosphere allowing realistic consideration of ethnic alternatives and options?

10.0 Does the curriculum promote values, attitudes, [illegible] and behaviors which support ethnic pluralism?

10.1 Does the curriculum help students examine differences within and among ethnic groups?

10.2 Does the curriculum foster attitudes supportive of cultural democracy and other democratic ideals and values?

10.3 Does the curriculum reflect ethnic pluralism?

10.4 Does the curriculum present ethnic pluralism as a vital societal force that encompasses both potential strength and potential conflict?

13.0 Is the multiethnic curriculum comprehensive in scope and sequence, presenting holistic views of ethnic groups, and an integral part of the total school curriculum?

13.1 Does the curriculum introduce students to the experiences of persons of widely varying backgrounds in the study of each ethnic group?

13.2 Does the curriculum discuss the successes and contributions of members of some group in terms of that group's values?

Guidelines 1971

2.1 Does the program focus on the social world as it actually is? R T C

2.2 Does the program emphasize pervasive and enduring social issues? R T

2.4 Does the program provide intensive and recurrent study of cultural, racial, religious, and ethnic groups? R T C

3.1 Does the program emphasize valid concepts, principles, and theories in the social sciences? R T

3.4 Does the program draw upon all of the social sciences and the history of the United States and the Western and non-Western worlds? R T

3.5 Does the program draw from what is appropriate in other related fields such as psychology, law, communications, and the humanities? R T

3.6 Does the program represent some balance between the immediate social environment of students and the larger social world? R T

3.7 Does the program include the study of man's achievements and those policies contrary to present national goals? R T

3.8 Does the program include a careful selection of that knowledge of most worth? R T

CHART G (Cont.)

14.0 Does the curriculum include the continuous study of the cultures, historical experiences, social realities, and existential conditions of ethnic groups with a variety of racial compositions?

14.1 Does the curriculum include study of several ethnic groups?

14.2 Does the curriculum include studies of both white and nonwhite groups?

14.3 Does the curriculum provide for continuity in the examination of aspects of experience affected by race?

17.1 Are the perspectives of different ethnic groups represented in the instructional program?

17.2 Are students taught why different ethnic groups often perceive the same historical event or contemporary situation differently?

18.0 Does the curriculum conceptualize and describe the development of the United States as a multidirectional society?

18.1 Does the curriculum view the territorial and cultural growth of the United States as flowing from several directions?

18.2 Does the curriculum include a parallel study of the various societies which developed in the geo-cultural United States?

19.3 Do students read and hear the poetry, short stories, novels, folklore, plays, essays, and autobiographies of a variety of ethnic groups?

19.4 Do students examine the music, art, architecture, and dance of a variety of ethnic groups?

19.5 Do students have available the artistic, musical, and literary expression of the local ethnic communities?

19.6 Are opportunities provided for students to develop their own artistic, literary, and musical expression?

20.0 Does the school foster the view of ethnic group languages as legitimate communication systems?

20.1 Are students taught about the nature of languages and dialects?

20.2 Is the student taught in his or her dominant language or dialect when needed?

20.3 Does the curriculum explore the role of languages and dialects in self-understanding and within and among ethnic groups?

20.4 Are the language policies and laws within the United States studied from political perspectives?

CHART H

Factor 8: Scope and Sequence-Skills

Guidelines 1976

8.0 Does the curriculum help students identify and understand the ever-present conflict between ideals and realities in human societies?

8.1 Does the curriculum help students identify and understand the value conflicts in problematic situations?

8.2 Does the curriculum examine differing views of ideals and realities among ethnic groups?

11.0 Does the curriculum help students develop decision-making abilities, social participation skills, and a sense of political efficacy needed for effective citizenship?

11.1 Does the curriculum help students develop the ability to distinguish facts from interpretations and opinions?

11.2 Does the curriculum help students develop skills in finding and processing information?

11.3 Does the curriculum help students develop sound knowledge, concepts, generalizations, and theories about issues related to ethnicity?

11.4 Does the curriculum help students develop sound methods of thinking about ethnic issues?

11.5 Does the curriculum help students develop skills in clarifying and justifying their values and relating them to their understanding of ethnicity?

11.6 Does the curriculum include opportunities to use knowledge, valuing, and thinking in decision-making on ethnic matters?

13.6 Does the curriculum provide for the development of progressively more complex concepts, abilities, and values?

16.0 Does the curriculum use comparative approaches in the study of ethnic groups and ethnicity?

16.1 Does the curriculum focus on the similarities and differences among ethnic groups?

16.2 Are matters examined from comparative perspectives with fairness to all?

17.0 Does the curriculum help students to view and interpret events, situations, and conflict from diverse ethnic perspectives and points of view?

17.3 Are the perspectives of each ethnic group presented as valid ways to perceive the past and the present?

Guidelines 1971

2.3 Does the program include analysis and attempts to formulate potential resolutions of present and controversial problems such as racism and war? R T

3.2 Does the program develop proficiency in methods of inquiry in the social sciences and in techniques for processing social data? R T O

3.3 Does the program develop students' ability to distinguish among empirical, logical, definitional, and normative propositions and problems? R T O

3.2 Do activities include formulating hypotheses and testing them by gathering and analyzing data? R T O

5.3 Do activities include the processes of making decisions about socio-civic affairs? R T O

7.1 Does the program help students organize their experiences? R O

7.2 Are learning experiences organized in such fashion that students learn how to continue to learn? R O

SECTION V

General Implications of The Guidelines for Large-Scale Assessment

The 1971 Guidelines provide direction for the evaluative process. The specific guidelines in 8.0 describe desirable charactertistics of assessment programs in general. Further, guidelines 4.0, while primarily concerned with objectives, has implications for desirable test characteristics. In fact, as we shall point out at the conclusion of this section, the 1971 Guidelines may even have implications for the types of tests which are most appropriate for student assessment in the social studies.

The 1971 Guidelines take an all too jaundiced view of paper-and-pencil tests. Of necessity the types of tests we are analyzing are of the paper-and-pencil variety. While we share concerns about a limited focus in the evaluation of student growth, we would not wish to see NCSS embrace a reactionary policy toward standardized testing of the type the NEA has manifested. This reactionary position, best articulated in the publication Alternatives to Standardized Testing (NEA, 1977), thoroughly misrepresents the potential utility of large-scale assessment programs.

The position which NCSS has taken toward evaluation in the general guideline 8.0 is in no way antithetical with standardized testing. For example, specific guideline 8.6 states:

> Evaluation data should offer students and teachers help in the course of learning and not merely at the conclusion of some marking period. (1971 Guidelines, p. 26)

The question may be asked, "What is the best method of methods for collecting what is essentially formative evaluation data?" The answer depends in part upon the purposes of testing and in selecting an instrument which is appropriate for the purposes. Clearly, a norm-referenced

test would be inappropriately used if the purpose of testing was to gather formative data. This example has implications for test publishers and test developers/users at the state and local levels. The appropriate and inappropriate uses of any test must be clearly communicated to potential test users. Test publishers must refrain from trying to give the impression that norm-referenced tests are appropriate for instructional decisions about individual students.

Many of the published tests we have analyzed are norm-referenced instruments, and as such are not directly congruent with general guideline 8.0. Norm-referenced paper-and-pencil tests should not be dismissed, however, as being incongruent with the 1971 Guidelines simply because they are not well matched with the general thrust of 8.0. Some norm-referenced tests are reasonably well matched with other guidelines where the purposes of evaluation are consistent with the intent of norm-referenced measurement.

While the implications of 8.0 are relatively limited for norm-referenced testing they certainly are not for criterion-referenced testing. The merits of criterion-referenced testing have been in part legitimized by the purported ability to provide information about what students can and cannot do; that is, to be diagnostic. While many test publishers maintain that they have criterion-referenced tests which allegedly possess the attributes delineated in 8.0, and endlessly pander their instruments consistent with the appropriate jargon, very few of their CRTs are equal to the task.

To this point in the report we have been somewhat cavalier about our use of certain terms. It may be wise for us to begin to be a bit more precise about what some of the terms mean to us. Initially, we will try and bring a modicum of sense to several testing terms; specifically

norm and criterion-referenced tests.

Norm-referenced tests (NRTs) are well suited for testing programs which have as their primary purpose the ranking or ordering of students by comparing them to a norm group. NRTs are composed of items which are not linked with any clearly described objectives. Items are selected for inclusion on an NRT in part by their ability to discriminate between students and also represent a wide range of difficulty. Therefore, norm-referenced tests are most appropriate for measuring the general achievement of groups of students. They are not good indicators of student attainment of specific objectives and have little utility for instructional planning on an individual basis. Of course, they are not intended for such uses.

If the purposes of testing include the diagnosis of individual student strengths and weaknesses and planning appropriate instructional intervention, then the use of criterion-referenced tests (CRTs) would appear to be appropriate. The qualification of the previous sentence is intentional because there seems to be some confusion as to what a CRT really is. For some, a CRT is any test which has a pre-established cut-score or performance standard. Equating the definition of a CRT only with the caveat that it possess a cut-score (a set percentage of items answered correctly on the total test) does not take into account the utilization of objectives or the consideration of test item characteristics. CRTs have also been interpreted as being tests which are composed of relatively clearly stated objectives with test items selected on the basis of their congruency with the objectives of interest. The number of items matched with the objectives is then determined (usually using some mystical process akin to capriciousness), as is the standard for "mastery" of the objectives. The most popular way of going about this task is to say; "we will have five items

measuring each objective, and we will consider a student as mastering the objective if he/she answers at least four of the items correct." This approach of loosely matching test items to objectives and arbitrarily selecting a cut-score (criterion for mastery) is the approach which most frequently passes for criterion-referenced testing.

It is our belief that a criterion-referenced test is a test which is used to describe "an individual's status with respect to a well-defined behavioral domain" (Popham, 1978). This definition of a CRT is not new: in fact it is essentially the same as was first put forward by Glaser (1963) fifteen years ago. Unfortunately, the original conception of criterion-referenced tests originated by Glaser and expanded upon by others (Popham and Husek, 1967) has gotten lost in the strenuous attempts of test publishers to beat the competition into the market place. We realize our acceptance of this definition of criterion-referenced tests is quite different from the popular perception of CRTs. However, this conception of criterion-referenced testing leads us and others to the conclusion that most tests in use today, purported to be CRTs, probably do not deserve that classification (Hambleton, 1978). The point we wish to make is that, of the tests which were reviewed during the course of this project, none could be considered a criterion-referenced test, and should be referred to as objectives-referenced tests.

The primary characteristic which a test must possess to truly be called "criterion-referenced" is that the domain of behaviors included on the test (sometimes called performance indicators) must be well defined. A domain, or set of behaviors, may be large or may be small, but its most important characteristic (for use in testing situations) is that it is well defined. The scheme which provides this definitional and descriptive

attribute is called test specifications. Test specifications may take any number of specific organizational forms, but any scheme for specifications includes the following components:

1. A general description - A statement that names and/or narratively describes the domain being assessed. It may include the conditions in which assessment will occur (e.g. given a list of positive and negative examples of the concept "prejudice", the student. . .), the format used for the test items, and content acceptable for inclusion in assessing the domain. The general description defines the domain of interest and places some general parameters upon it.

2. A Sample Test Item - The purpose of including a sample test item is to illustrate how test items used to measure the domain will look. All items used on a test which assess a single domain will be of the same format as the sample item. The sample item provides, at a glance, much of the interest of the measurement process, what the students will be doing, and how they will be doing it.

3. Stimulus Attributes - All stimulus materials which precede the actual selection or construction of a response by a test-taker are described by the stimulus attributes. The content and format of the directions are included in this section. Any stimulus materials are illustrated as are the rules for their inclusion (e.g. for reading passages—minimum and maximum sentence length, type and amount of acceptable

punctuation, readability level, etc).

4. Response Attributes - The types and characteristics of responses made to the test item stimulus are called the response attributes. For example: whether the type of response is selected or constructed, acceptable number of response alternatives in selected response options, characteristics of distractors and the criterion for correctness of constructed responses. The format and appearance of the responses are described, where appropriate.

5. Supplement - If the previous four sections are not entirely descriptive of acceptable content of format, a supplement may be included to better describe the assessment process. For example, if the domain of interest includes the definition of economic terms it would be too cumbersome to include all acceptable terms in the general description. Therefore, a supplement would be added to delineate terms acceptable for testing. Another example of what may be included in the supplement would be the formulae used to compute readability. This may be particularly important since the different formulae have the rather disconcerting characteristic of providing different estimates of the readability level of the same passage.

The generation of test specifications accomplishes several important ends:

1. Communicating to the public (usually in a less technical format) what the schools are testing and how it is tested.

2. Communicating to instructional personnel (and perhaps students) the outcomes measured by the test, how they are measured, and perhaps most importantly, the types of possible systematic errors students may make. This final characteristic, coupled with a systematic reporting of results, provides the most significant potential for more precise diagnostic testing.

3. Communicating to persons charged with writing test items acceptable item characteristics.

4. The ability to utilize appropriate statistical methods for estimating a person's status relative to the behavioral domain of interest.

The emphasis upon describing and defining the behaviors of interest has implications for test developers, both conceptually and in practical situations such as item development. The descriptive scheme which we are presenting here is also consistent with the 1971 specific guidelines 4.1, 4.3, 4.4, and 4.5:

> 4.1 Are objectives carefully selected and formulated?
>
> 4.3 Are general statements of goals translated into specific objectives conceived in terms of behavior and content?
>
> 4.4 Are classroom instruction and materials based upon clearly stated objectives?
>
> 4.5 Does classroom instruction enable students to see their goals clearly in brief instructional sequences and lengthy units of study?

These specific guidelines were written in part as a result of the behavioral objectives movement and are intended primarily for instructional planning at the classroom level. With a minor shift in

emphasis they are also appropriate for describing a criterion-referenced test of the type we are advocating. For example, very narrow behavioral objectives may present too limited a focus to represent a domain of behaviors. Keeping these specific guidelines in mind, CRT publishers/developers would do well to inform test users of:

1. The objectives (preferably for us, domains) included on the tests.

2. The process used to review and select objectives included on their criterion-referenced tests.

3. Test specifications.

4. Reasonable instructional support systems based not only upon right/wrong responses but systematic wrong-response patterns.

SECTION V

Development of the Test Review Process

The review form (page 46) developed for use in analyzing published social studies tests represents our best efforts to maximize the amount of information we could obtain in a reasonable amount of time. It is by no means perfect, some may feel it is not particularly fair. But it does represent a first, and we certainly hope not the last, attempt to objectively evaluate testing instruments.

The form is divided into two major areas, Scope and Sequence-Content and Scope and Sequence-Skills. This breakdown was congruent with the analysis of the Guidelines for their appropriateness for large-scale assessment (see criteria pp 31-33 in section IV). The two factors above were the ones which appeared to be the most general grouping of the

Guidelines we could infer after deletion of the general and specific guidelines judged as inappropriate for large-scale assessment. This should in no way be construed as a value judgment by us about the relative importance of any of the general or specific guidelines.

As the result of numerous discussions, the number of tests analyzed by the Committee on Testing was limited. The COT first considered what to do about published tests that were designed for single, traditional disciplines—e.g. economics or sociology. The COT decided to eliminate special subject tests from the list because of time constraints. In addition, the COT judged that these tests were less appropriate to the criteria from the guidelines.

Secondly, the COT eliminated non-content specific psychological tests on the basis that these tests are too numerous and too diverse in the focus of assessment qualities and more appropriate for assessment of local curriculum goals. Furthermore, these tests tend to center on factors of self-worth (ego), which the COT had earlier determined was more appropriately the domain of local school assessment.

Finally, the COT chose to examine only those sections of each test designated "social studies". Frequently the skills and sometimes content, referred to in the guidelines could be found under language arts, mathematics and/or science. Since scores on the tests are not reported by skill function or specific content (e.g. pollution), analyses of the other than social studies sections would provide little knowledge of value.

The COT proceeded to analyze the social studies tests in the following stages:

FIRST, individual members rated a small sample of the tests and then determined the reliability of the ratings.

Form/Level _____ Total Items _____

I. Scope and Sequence-Content

A. Item Content (Stimulus)

Breadth Across Disciplines

	Item Numbers	Total
Economics (use of resources) (consumerism)		
Geography		
Am. History		
W. History		
Soc./Anth.		
Govt.		
Affect		
Law		
Other (careers)		
Total Number of Items		

Pervasive Social Issues

	Item Numbers	Total
War		
Human Rights (discrimination) (rel. conflict) (racism)		
Conservation (pollution) (population)		
Urbanism		
Soc. Welfare		
Income Dist.		
Total (PSI)		
Ratio with Total ___ = ___ %		Total

Ethnic Content	Ratio with Total	% Total
Ethnic Questions		
Ethnic Symbols/Heroes (No.) (tokenism)		
Ethnic Concepts (No.)		

Source of Ethnicity

Black		
Latino		
Asian		
Native Am.		
Jewish		
European (other)		

B. Response Processes

Simple Recall _____

_____ Number _____ % Total _____

Other _____ Number _____ % Total

II. Scope and Sequence-Skills

A. Distribution of Pictorial/Reference Skills

	Item Numbers	Total
Tables	_____	_____
Location (Maps, Globes, Models)	_____	_____
Graphs (Symbols)	_____	_____
Visuals (Pictures, Cartoons)	_____	_____
Reference Materials (shelf reference)	_____	_____
Chronology (Timelines, Sequences)	_____	_____
	Total	

B. Information Processing Skills

Inferencing behavior	_____	_____
Distinguish fact/opinion	_____	_____
Relevant/irrelevant information	_____	_____
Classification	_____	_____
Compare/contrast	_____	_____
Hypothesis generation/prediction	_____	_____
Hypothesis testing	_____	_____
Forming generalizations	_____	_____
Evaluate sources of information	_____	_____
Interpretation	_____	_____
Problem Identification	_____	_____
Other	_____	_____

C. Evaluation

Judges worth of decisions/values	_____	_____

Ratio A/B =

SECOND, the Committee discussed differences in the use of terms that seemed to confuse the rating process. On the basis of these discussions minor revisions were made in the test review form.

THIRD, the Committee assigned a sample of the tests to each individual and the full committee received the ratings on each test.

SECTION VII

Results

The following reporting of results has been done in a narrative format. We have selected this format for several reasons. The data which we have collected and analyzed does not lend itself to strict quantitative analysis. Tests of significance, in this case, would have no meaningfulness. This is in part due to the nature of the instrument which we developed for reviewing the test. It is subjective and judgmental reflecting our own untested beliefs and biases. Further, no quantitative attempts were made to evaluate inter-rater agreement, given the amorphous nature of the jargon and the eclectic nature of the task, any such attempts would have been counterproductive. With a sufficient staff of trained raters this may be possible to do. With limited resources it is almost impossible. As a result, we have attempted to summarize the more meaningful findings in our narrative. A more delicate analysis of this type of data will have to wait upon more diligent and resourceful investigators.

Commercial Tests. As one would expect, many of the tests contained a representative number of items from the major disciplines. A general

trend seems to be emerging in the more recently published tests toward greater emphasis in the area of economics. Particularly, grades 6-9 seem to be the major focus of this increased attention. This can be seen in the relatively lower percent of items dealing with economics on the upper-level tests, e.g. TAP, STEP (J/10-12), CTBS (S/4) and Metropolitan (JS-ADV 1). A notable exception is the prototype of the SRA Achievement Series (1/H). Probably consistent with nationwide curriculum trends is the emphasis upon geography in the lower grades. Generally, American History still retains a prominent position on several of the upper-level tests, with the CTBS (S/4) and the Metropolitan Achievement Test (JS-ADV/2) as exceptions. Items dealing with sociology seem to be well represented, although many of these items may also have been classified as World History. These items in many cases dealt with non-Western culture and customs. For a few tests, greater than 20% of their items were placed in the "other" category. On the STEP (H/6-7), items dealing with an unidentified primitive culture and items dealing with farming techniques were classified as "other". On the Metropolitan (JS/ELEM) items dealing with changing personal physical characteristics, unidentified primitive people, planting trees in parks, and selling animals in a pet store were also classified as "other" (see Table 6).

Human rights and conservation generally received the most attention in terms of pervasive social issues. On one test (STEP-J/10-12), social welfare issues were used as stimulus material. STEP J/10-12 seemed to be the test which could be considered as having the highest issue orientation, and included such issues as housing trends, the position of women, labor management issues, and cultural change. Issues such as

war, urbanism, social welfare, and income distribution received relatively little attention (see Table 7).

The ethnic content of the tests reviewed generally dealt more with cultural concepts in a significant fashion rather than in a token fashion. There were some exceptions to this general trend. While tokenism is not the general trend for the ethnic content of the tests, virtually all tests had a rather low overall proportion of items devoted to various ethnic groups. Blacks, Latin-Americans, and Native Americans were the most represented ethnic groups (see Table 8).

In the broadest breakdown classification of cognitive skills which we analyzed (simple recall vs. higher thought processes) the results of the analysis show little discernable pattern. Some tests are heavily recall (ITED/X5) with as much as 97% of their items falling into this classification. Others, such as the STEP tests (forms F,H, and J), are heavily weighted toward higher-order thought processes (88, 80 and 82%) respectively (see Table 9). In terms of visual stimuli for test items most are locational (maps), graphs, and visuals (pictures, cartoons). There is little use of reference materials or time-lines as stimulus materials (see Table 9).

Our attempts at identifying, in a more precise fashion, the specific higher-order cognitive processes are found in Table 10 . The two higher-order skills which are found most frequently are those dealing with interpreting and inferencing behaviors. The most noteworthy of the other higher-order skills were evaluating sources of information and distinguishing between fact and opinion statements. While these latter two skills do receive attention on a few tests, they are by no means found frequently when the tests are examined overall. A number

TABLE 6
DISTRIBUTION OF TEST ITEMS ON
NORM-REFERENCED SOCIAL STUDIES TESTS
BY CONTENT AREAS (IN PERCENT)

Test Name	Economics	Geography	American History	World History	Sociology/ Anthropology	Government	Law	Career Educ.	Affect	Other	Total # of Test Items
EDUCATIONAL TESTING SERVICE											
Step (F/4-5) (P)	22	20	4	14	14				6	20	50
Step (H/6-7) (P)	30	8	6		14	2			2	38	50
Step (J/10-12 (P)	14	8	42		22	8				6	50
CTB McGRAW-HILL											
CTBS (S/1 Gr. 3) W.Va.		46	16		19	3	16				37
CTBS (S/3 Gr. 9) W.Va.	55	3	13		15	3				13	40
CTBS (S/4 Gr. 11) W.Va.	13		13	15	46	13					39
CTBS (S/3 Gr. 9)	43	3	13		18	10				15	40
PSYCHOLOGICAL CORPORATION											
Metro. Ach. (JS-ADV 1) (P)	12	22	14	8	12	28				4	50
Metro. Ach. (JS-ELEM) (P)	2	20	20		2	4	9		9	33	45
HOUGHTON-MIFFLIN											
TEST ACAD. PROG. (S/9-12)	10	10	25	13	7	18				17	120*
SCIENCE RESEARCH ASSOCIATES											
SRA Ach. (1/E) 4-6 (P)	13	63	18							8	40
SRA Assess. (E/BLUE) 4-6	20	30	28	7	10	5					40
SRA Ach. (1/F) 6-8 (P)	18	7	7	10	23	15				20	40
ITED (X5/9-12)	23		20	10	7	16	13			10	30
SRA Ach. (1/H) 9-12 (P)	32	5	38		5	17				3	40
HARCOURT BRACE JOVANOVICH											
Stanford Ach. (A/ INT 2)	7	20	17	7	13	11				24	54
Stanford Ach. (X/H.S.)	6	21	41	22		10					63

TABLE 7
DISTRIBUTION OF TEST ITEMS ON
NORM-REFERENCED SOCIAL STUDIES TESTS
BY ISSUE ORIENTATION

Test Name		War	Human Rights	Conservation	Urbanism	Social Welfare	Income Dist.	Other	Total # of Test Items
EDUCATIONAL TESTING SERVICE									
Step (F/4-5)	(P)		2	2	2			8	50
Step (H/6-7)	(P)		8	10	6				50
Step (J/10-12)	(P)		28	16		20		22	50
CTB McGRAW-HILL									
CTBS (S/1 Gr. 3) W.Va.									37
CTBS (S/3 Gr. 9) W.Va.		10	15						40
CTBS (S/4 Gr. 11) W.Va.									39
CTBS (S/3 Gr. 9)			10	30					40
PSYCHOLOCIAL CORPORATION									
Metro. Ach. (JS-ADV 1)	(P)						2		50
Metro. Ach. (JS-ELEM)	(P)								50
HOUGHTON-MIFLIN									
TEST ACAD. PROG. (S/9-12)			3						120
SCIENCE RESEARCH ASSOCIATES									
SRA Ach. (1/E) 4-6	(P)			10	10				40
SRA Assess. (E/BLUE) 4-6						3			40
SRA Ach. (1/F) 6-8	(P)		5	10				10	40
ITED (X5/9-12)			3					3	30
SRA Ach. (1/H) 9-12	(P)			3					40
HARCOURT BRACE JOVANOVICH									
Stanford Ach. (A/INT 2)				2					54
Stanford Ach. (X/H.S.)									63

TABLE 8
DISTRIBUTION OF TEST ITEMS ON
NORM-REFERENCED SOCIAL STUDIES TESTS
BY ETHNIC CONTENT AND SOURCE OF ETHNICITY

Test Name	Ethnic Content		Source of Ethnicity						Total # of Test Items
	Token.	Conc.	Black	Latin Am.	Asian	Nat. Am.	Jew.	Eur.	
EDUCATIONAL TESTING SERVICE									
Step (F/4-5) (P)	4	10		4	2			8[a]	50
Step (H/6-7) (P)	6	6	8	2	2				50
Step (J/10-12) (P)	16	6	4		6	12			50
CTB McGRAW-HILL									
CTBS (S/1 Gr. 3) W.Va.		14				14			37
CTBS (s/3 Gr. 9) W.Va.		10	10						40
CTBS (S/4 Gr. 11) W.Va.	3	15	18						39
CTBS (S/3 Gr. 9)		10	10						40
PSYCHOLOGICAL CORPORATION									
Metro (ADV 1) (P)									50
Metro. (ELEM) (P)	2	9			2	2		7	45
HOUGHTON-MIFLIN									
TEST ACAD PROG. (S/9-12)	.8	1.6	1.6					.8	120*
SCIENCE RESEARCH ASSOCIATES									
SRA Ach. (1/E) 4-6 (P)		5			5				40
SRA Assess. (E/BLUE) 4-6	8		3		3	7.5			40
SRA Ach. (1/F) 6-8 (P)		28	7.5			20			40
ITED (X5/9-12)	3		3						30
SRA Ach. (1/H) 9-12 (P)									40
HARCOURT BRACE JOVANOVICH									
Stanford Ach. (A/INT 2)	2		2						54
Stanford (X/H.S.)		3		2	2				63

[a]NOTE: These percentages reflect the categories that were not immediately recognizable for their other content. i.e. primitive people with no clues as to their origin or location.

TABLE 9
DISTRIBUTION OF TEST ITEMS ON
NORM-REFERENCED SOCIAL STUDIES TESTS
BY RECALL VS. HIGHER COGNITIVE SKILLS AND PICTORIAL/REFERENCED SKILLS

Test Name	Skill Level		Pictorial/Reference Skills						Total # of Test Items
	Recall	Higher	Tables	Iteration	Graphs	Visuals	Refer.	Chron.	
EDUCATIONAL TESTING SERVICE									
Step (F/4-5) (P)	12	88		12	16	42		4	50
Step (H/6-7) (P)	20	80		8	20	26			50
Step (J/10-12) (P)	18	82		10	24	4			50
CTB McGRAW-HILL									
CTBS (S/1 Gr. 3) W.Va.	92	8		14	11	3			37
CTBS (S/3 Gr. 9) W.Va.	33	67		3	15				40
CTBS (S/4 Gr. 11) W.Va.	54	46		10	15		3	3	39
CTBS (S/3 Gr. 9)	43	57		3	17				40
PSYCHOLOGICAL CORPORATION									
Metro. (ADV 1) (P)	73	27	20	10				2	50
Metro. (ELEM) (P)	38	62	16	20		9			45
HOUGHTON-MIFLIN									
TEST ACAD. PROG. (S/9-12)	59	41	3	8	16				120*
SCIENCE RESEARCH ASSOCIATES									
SRA Ach. (1/E) 4-6 (P)	45	55	8	10		40	5		40
SRA Assess. (E/BLUE) 4-6	88	12		20	15	15			40
SRA Ach. (1/F) 6-8 (P)	33	67		15	10	65			40
ITED (X5/9-12)	97	3							30
SRA Ach. (1/H) 9-12 (P)	75	25			8	30			40
HARCOURT BRACE JOVANOVICH									
Stanford Ach. (A/INT 2)	59	41		19	11	9	11		54
Stanford Ach. (X/H.S.)	71	29		22	8			6	63

TABLE 10
DISTRIBUTION OF TEST ITEMS ON
NORM-REFERENCED SOCIAL STUDIES TESTS
BY INFORMATION PROCESSING SKILLS

Test Name	Inference	Fact vs. Opinion	Relevance v. Irrelevance	Classif.	Compare/ Contrast	Hypoth. Generation	Hypoth. Testing	Form General.	Evaluation Sources	Interp.	Problem Ident.	Judges Worth	Total # of Test Items
EDUCATIONAL TESTING SERVICE													
Step (F/4-5) (P)	18	2			2	6		2	2	32	2	4	12
Step (H/6-7) (P)	24	2	6			8	2		8	24	4		2
Step (J/10-12) (P)	22		6		6	2	2	4	12	24		4	2
CTB McGRAW-HILL													
CTBS(S/1 Gr. 3)W.Va.		3				2					3		
CTBS(S/3 Gr. 9)W.Va.	3	15				13			.15	20	3		
CTBS(S/4Gr. 11)W.Va.		5	5	10	3	10	3		3	7			
CTBS(S/3 Gr. 9)	7	15	2	5					15	10	3		
PSYCHOLOGICAL CORPORATION													
Metro. (ADV 1) (P)	12			4	2		4		2	2			2
Metro. (ELEM((P)	2				7	2	2	7		29			
HOUGHTON-MIFLIN													
TEST ACAD. PROG.	13		.8	3	9	.8	1.6	3		6		7	7
SCIENCE RESEARCH ASSOCIATES													
SRA Ach.(1/E)4-6 (P)				3		18	3	10		20			
SRA Assess. 4-6			3				3			7			
SRA Ach.(1/F)6-8 (P)	23				5	8	2	7		20	2.5		
ITED (X5/9-12)	3												
SRA Ach.(1/H)9-12(P)	5					2.5	2.5			13	2.5		
HARCOURT BRACE JOVANOVICH													
Stanford Ach.(A)	15			2	2		4		4	15			
Stanford Ach.(X)	2			5	8	6		6		2			

of the skills receive virtually no apparent attention, e.g. problem identification and evaluating the worth of decisions/values.

State Testing Programs. The Committee on Testing early decided that the original task before the Committee should be broadened (see page 4) to include analyses of those social studies tests utilized by state agencies and/or those commercial tests uniquely adapted to meet individual state objectives. Letters were sent to a CS4 member in each state asking them to provide the following information (see letters in appendix):

1) A list of statewide social studies objectives (K-12)—both affective and cognitive.

2) Names and copies of social studies tests used in state testing programs—both norm-referenced and/or criterion-referenced.

3) Changes that are underway or contemplated in the testing of social studies outcomes.

A description of the materials collected appears in the appendix. (The COT contemplates conducting a more precise survey in the near future).

When the survey was first planned, it was assumed by the COT that the classic curriculum model would be found in the several states. That is first, that each state surveyed would have a carefully worded description or definition of social studies curriculum for grades K-12. Secondly, that each state or locality would exhibit carefully developed curriculum/instruction guidelines from which instructional objectives could be prepared. Finally, it was assumed that measurement instruments of some type were being used or developed to evaluate the curriculum/instruction program in social studies.

Results of our survey, with forty-one(41) of the fifty (50) states reporting, suggest findings as follows:

(1) Twenty-three (23) states report that their state has prepared a document outlining statewide goals of the social studies curriculum. These documents range from carefully defined philosophies of social studies to brief and broad statements about humankind. Two states report that the NCSS Guidelines were adopted.

(2) Fourteen (14) states have written or are writing curriculum/instruction objectives as mandated guides for some aspect of the social studies programs in elementary and secondary schools.

(3) Twenty-three (23) states report that some type of competency based program is being mandated that directly or collaterally impinges upon the content of the social studies curriculum. The focus of the CBE movement in social studies is upon citizenship education and economic education (frequently called free enterprise).

(4) Eighteen (18) states have statewide testing programs that are specifically related to one or more social studies subjects—tests that were designed to be unique to that state. Usually the testing occurs twice in the elementary grades (K-7) and/or twice in the secondary years (7-12).

(5) In the states with statewide testing programs, four (4) utilize achievement (NRT) tests; two (2) use criterion-referenced tests and twelve (12) use exercises from the National Assessment of Educational Progress.

It is clear that approximately half of the states will soon require competency testing programs that impinge on some aspect of the social studies curriculum. Yet these states will, in the main, conduct competency testing without previously determined curriculum objectives. Thus, we will have a social studies curriculum by "deflection" from the measures of performance. It is not at all clear that the NCSS Curriculum Guidelines (1971 and 1976) have significantly influenced curriculum trends at the state level.

The responses from many states convey a definite feeling that the state should not mandate social studies curriculum. This explains, at least in part, why the definition of social studies is so wide ranging and ambiguous. It clearly required that the Committee reject the assumption that the classic curriculum model exists in state social studies programs.[1]

The Committee received tests from eighteen (18) states. Tests from ten (10) states were analyzed by the COT. The results of these analyses are reported in Tables 11 through 15. Tests from the other eight (8) states were repetitious tests circulated by commercial companies under the heading adapted for a particular state. Upon examination, we found that the social studies section of these tests were identical to the original test form.

Summary data on the Scope and Sequence-Content included on these tests were reported in Table 11 and Table 12 . The following conclusions can be reported:

---

[1]Such states as New York and Michigan are notably unique--they are models of a clearly planned curriculum.

1) State Achievement tests (NRT) tend to sample the content of traditional disciplines of economics, geography, American History and government. World history was a notable exception. Sociology was included extensively only on the one achievement test at the 12th grade level. On the other hand, achievement tests tended not to include related subjects—law, career education, etc.

2) Criterion (so-called) and/or competency tests tended to focus upon special social studies contents—usually law, careers, citizenship, etc. Fewer items from traditional disciplines were included.

3) The state assessments of educational progress broadly sampled the traditional disciplines along with items from careers, law, etc. Unclassified items on these instruments tended to examine personal development and respect for others. (NOTE: the quality of exercises on these tests was highly variant.)

4) Pervasive issues—content deemed important in the Guidelines—were not often found on the achievement and criterion tests. There were more items found on the NAEP type instruments, but rarely did these items account for more than 10% of the test.

5) The most pervasive issue represented on the tests was the issue of human rights—civil and religious rights.

6) State social studies tests do not contain signigicant numbers of items that represent ethnic pluralism.

While tokenism (e.g. hero reference) were not largely present, the number of ethnic concepts represented in the stimuli and/or responses on items were severely limited. The National Assessment of Educational Progress (NAEP) type items were generally superior.

Summary data on the Scope and Sequence-Skills inclusion on state social studies tests were reported in Table 13 and 15 . The following conclusions can be reported:

1) Recall versus higher-order skill items were highly variant. There was no consistent pattern by test type.

2) Basic skills of information gathering and reporting were present on all types of state tests—but interestingly enough the criterion/competency reference tests did the least to test pictorial/reference skills. NAEP type instruments seemed slightly superior on these dimensions.

3) The NAEP type exercises seemed to broadly sample the thinking skills—somewhat more so than achievement and competency tests. The greatest emphasis would appear to be skills of interpretation, hypothesis generation, inference and compare and contrast.

While the sample of state developed tests was small, some trends seem notable:

1) States that have required testing programs tend to use the social studies section of commercial tests to report scores to schools and individuals. These tests tend to sample traditional content areas, slighting social issues and ethnic pluralism, but broadly sample informational

TABLE 11
DISTRIBUTION OF TEST ITEMS ON
STATE SOCIAL STUDIES TESTS
BY CONTENT AREAS (In Percent)

State Test Name— Grouped By Type of Test		Economics	Geography	American History	World History	Sociology/ Anthropology	Government	Law	Career Educ.	Affect	Other	Total # of Test Items
ACHIEVEMENT TESTS												
N.Y. State Ach.	(Gr. 11)	19	17	28			28				9	47
N.Y. Regents	(Gr. 11)	22	6	14	10	8	22				6	49
Minn. H.S. Ach.	(Gr. 7)			100								117
	(Gr. 8)		100									108
	(Gr. 9)						88				12	88
	(Gr. 9)			100								128
	(Gr. 11)			100								113
	(Gr. 12)	18		6	1	45	31					110
SCORE (N.C.)	(Gr. 3)	26	54	4			15				2	54
	(Gr. 6)	21	25	6	3	2	22			3	19	68
CRITERION/COMPETENCY TESTS												
R.I. Criterion	(CRT #21)	11	9				37	30			13	46
	(CRT # 1)	41			4	4	13		24		13	46
Penn. Law Comp.	(Prim.)							100				75
	(Inter.)							100				66
	(Sec.)							100				100
ASSESSMENTS OF EDUCATIONAL PROGRESS												
Ill. AEP	(Gr. 8)						100					28
	(Gr. 11)						100					30
Del. AEP	(Gr. 4)	13	53		3	27	3					30
	(Gr. 8)	14	26	23	14	14	3			6		35
Mich. AEP	(Gr. 4)		22	6					22		50	113
	(Gr. 7)	14	15	21	3	1	13	9	1		22	183
Mass. AEP	(Age 9)	29	18	12			21	3			18	34
	(Age 17)	37	15	12	5		29	2				41

TABLE 12
DISTRIBUTION OF TEST ITEMS ON
STATE SOCIAL STUDIES TESTS
BY ISSUE ORIENTATION (In Percent)

State Test Name— Grouped By Type of Test		War	Human Rights	Conser-vation	Urbanism	Social Welfare	Income Dist.	Other	Total # of Test Items
ACHIEVEMENT TESTS									
N.Y. State Ach.	(Gr. 11)		2	9	9				47
N.Y. Regents	(Gr. 11)		4						49
Minn. H.S. Ach.	(Gr. 7)								117
	(Gr. 8)		1						108
	(Gr. 9)								88
	(Gr. 9)								128
	(Gr. 11)								113
	(Gr. 12)		3						110
SCORE (N.C.)	(Gr. 3)	2							54
	(Gr. 6)	2	4						68
CRITERION/COMPETENCY TESTS									
R.I. Criterion	(CRT #21)		2		4				46
	(CRT # 1)								46
Penn. Law Comp.	(Prim.)								75
	(Inter.)								66
	(Sec.)								100
ASSESSMENTS OF EDUCATIONAL PROGRESS									
Ill. AEP	(Gr. 8)		11						28
	(Gr. 11)		67						30
Del. AEP	(Gr. 4)		17	7					30
	(Gr. 8)		1	9					35
Mich. AEP	(Gr. 4)								113
	(Gr. 7)	3	6	6	2				183
Mass. AEP	(Age 9)		3	3					34
	(Age 17)	2	2			2			41

TABLE 13
DISTRIBUTION OF TEST ITEMS ON
STATE SOCIAL STUDIES TESTS
BY RECALL vs. HIGHER COGNITIVE SKILLS AND
PICTORIAL/REFERENCE SKILLS (In Percent)

State Test Name— Grouped By Type of Test		Skill Level		Pictorial/Reference Skills						Total # of Test Items
		Recall	Higher	Tables	Location	Graphs	Visuals	Refer.	Chron.	
ACHIEVEMENT TESTS										
N.Y. Ach.	(Gr. 11)	45	55		2	13	9			47
N.Y. Reg.	(Gr. 11)	53	47	6	8		6			49
Minn. Ach.	(Gr. 7)	100								117
	(Gr. 8)	100								108
	(Gr. 9)	88	12		3			2		88
	(Gr. 9)	100								128
	(Gr. 11)	100								113
	(Gr. 12)	100								110
SCORE N.C.	(Gr. 3)	69	31		13		22		4	54
	(Gr. 6)	55	45	19	2				3	68
CRITERION/COMPETENCY TESTS										
R.I.	(CRT #21)	48	52		9		9			46
	(CRT # 1)	48	52	13		9				46
Pa. Law	(Prim.)	100								75
	(Inter.)	100								66
	(Sec.)	100								100
ASSESSMENT OF EDUCATIONAL PROGRESS										
Ill.	(Gr. 8)	100								28
	(Gr. 11)	100								30
Del.	(Gr. 4)	57	43							30
	(Gr. 8)	40	60		14	14	6	3	6	35
Mich.	(Gr. 4)	48	52	5			11			113
	(Gr. 7)	17	83							183
Mass.	(Age 9)	70	30					6		34
	(Age 17)	60	40	2	2	7	12	2		41

TABLE 14
DISTRIBUTION OF TEST ITEMS ON
STATE SOCIAL STUDIES TESTS
BY ETHNIC CONTENT AND SOURCE OF ETHNICITY (In Percent)

State Test Name— Grouped By Type of Test		Ethnic Content		Source of Ethnicity						Total # of Test Items
		Token.	Conc.	Black	Latin Am.	Asian	Nat. Am.	Jew.	Eur.	
ACHIEVEMENT TESTS										
N.Y. Ach.	(Gr. 11)		9	2					6	47
N.Y. Reg.	(Gr. 11)		4	4						49
Minn. Ach.	(Gr. 7)									117
	(Gr. 8)									108
	(Gr. 9)									88
	(Gr. 9)									128
	(Gr. 11)									113
	(Gr. 12)	1	6	3	1		1		1	110
SCORE N.C.	(Gr. 3)									54
	(Gr. 6)		2		2					68
CRITERION/COMPETENCY TESTS										
R.I.	(CRT #21)		2						2	46
	(CRT # 1)		9	4	4	4				46
Pa. Law	(Prim.)									75
	(Inter.)									66
	(Sec.)									
ASSESSMENT OF EDUCATIONAL PROGRESS										
Ill.	(Gr. 8)									28
	(Gr. 11)									30
Del.	(Gr. 4)		10	7				3		30
	(Gr. 8)	3	31	6	3	14	9		9	35
Mich.	(Gr. 4)									113
	(Gr. 7)		7	2		2		1	2	183
Mass.	(Age 9)	3		3						34
	(Age 17)									41

TABLE 15
DISTRIBUTION OF TEST ITEMS ON
STATE SOCIAL STUDIES TESTS
BY INFORMATION PROCESSING SKILLS (In Percent)

State Test Name—Grouped By Type of Test	Inference	Fact vs. Opinion	Relevance v. Irrelevance	Classif.	Compare/ Contrast	Hypoth. Generation	Hypoth. Testing	Form General.	Evaluation Sources	Interp.	Problem Ident.	Judges Worth	Total # of Test Items
ACHIEVEMENT TESTS													
N.Y. Ach. (Gr. 11)	2	2				2		4		45			47
N.Y. Reg. (Gr. 11)				2	2	8		10		14			49
Minn. Ach. (Gr. 7)													117
(Gr. 8)													108
(Gr. 9)		2	1				1		1				88
(Gr. 9)													128
(Gr. 11)													113
(Gr. 12)													110
SCORE N.C. (Gr. 3)	9			2	4	6			4	7			54
(Gr. 6)	3			6	3	7	3	2	4	10	6		68
CRITERION/COMPETENCY TESTS													
R.I. (CRT #21)	13					4			26	9			46
(CRT # 1)	4				15			4	4	22			46
Pa. Law (Prim.)													75
(Inter.)													66
(Sec.)													100
ASSESSMENT OF EDUCATIONAL PROGRESS													
Ill. (Gr. 8)													28
(Gr. 11)													30
Del. (Gr. 4)					17	13		7	3	3			30
(Gr. 8)	6		6		3	17		9	6	14			35
Mich. (Gr. 4)				11	17	15				11	4		113
(Gr. 7)	9		3	5	13	25		13	1	13	2	1	183
Mass. (Age 9)		6			3			3		12			34
(Age 17)	5	2				10		5	2	15			41

and thinking skills.

2) Achievement tests produced by commercial publishers for a state are likely to contain the same qualities as under one above.

3. Competency and so-called criterion-referenced tests appear to be knowledge based, sampling content areas of recent political concern--consumer economics, citizenship, law, etc. While these tests may be targeted to selected objectives, the objectives appear simplistic.

4) The influence of the National Assessment of Educational Progress movement is clearly evident in the state programs. Several states have conducted their own assessment and this report contains analyses of four NAEP tests currently in use. The exercises on the NAEP type instruments vary in the quality of stimulus and/or responses. The Michigan program provides the best examples of sound test development in the states.

SECTION VIII

Conclusions and Recommendations

It was the responsibility of the COT to 1) derive classes of learning outcomes based upon the 1971 and 1976 Guidelines and 2) to analyze commercially produced tests for congruency with those learning outcomes. In order to accomplish this task, the COT was forced to make a series of assumptions:

1. The NCSS Curriculum Guidelines (1971 and 1976) represent a definition of the social studies that can be

translated into intended learning outcomes.

2. Social studies curricula (and instructional plans) can be developed by states and localities that incorporate the Guidelines.

3. Publishers of social studies tests can determine learning outcomes from the Guidelines and construct instruments that evaluate those outcomes.

As reported in Section IV, the COT identified eight (8) factors that were present in the Guidelines of which only two factors contained guideline statements that referenced learning outcomes appropriate for large-scale assessment--Scope and Sequence-Content and Scope and Sequence-Skills. Two additional factors--self-worth and student participation--contained guideline statements referencing student behaviors of interest to instructional planners but not appropriate for large-scale assessment.

As the COT examined the Scope and Sequence-Content factor, it became clear that the Guidelines provided broad criteria for describing the content areas of the social studies--traditional disciplines, pervasive social issues and ethnic pluralism. The Guideline statements did not provide insights into the social studies as a discipline nor to the sequencing of social studies concepts.

The Scope and Sequence-Skills factor contained guideline statements that frequently confused rather than defined social studies behaviors. Operational differences between inquiry, reflective thinking, valuing, decision-making and critical thinking were not explained nor were their components described. Even though Guideline 13.6 (Multiethnic 1976) insists that the social studies curriculum should provide for a hierarchy of skill development, the Guidelines do not define this hierarchy. The COT was

forced to examine such publications as Gross and Chapin, Teaching Social Studies Skills and Bloom, Taxonomy of Educational Objectives for help in constructing the Test Review form. Essentially, we found that the NCSS Guidelines do not appropriately define student skill behaviors in social studies that would assist publishers in test development, educational leaders in curriculum development and the COT in the completion of its task. Therefore, all of the conclusions that follow in this section are tempered by the knowledge that the NCSS Guidelines do not describe skills essential or unique to social studies.

Our evaluation of published social studies tests has pointed up both deficiencies and strengths relative to the 1971 and 1976 Guidelines. First, the bitter pill and afterwards the kudos.

First, in the general grouping of the 1971 Guidelines that we have called Scope and Sequence-Content (2.1-2.7), (3.1-3.8), it would appear to the COT that there are some mis-matches between these specific guidelines and published tests. Specifically, these guidelines call for attention to be drawn in the curriculum (for us tests) toward pervasive social issues. Discounting the notion that test publishers can divine what the "real" social world is better than anyone else (2.1, 1971), closure on the identification of pervasive social issues may be relatively easy to reach. Both the 1971 and 1976 Guidelines see at least one major component of these issues as consisting of the experiences of different ethnic groups. While test publishers have apparently attempted to embrace the intent of specific objective 7.7 (1976) and avoid tokenism, the published tests by no means reflect an honest examination of the following 1976 specific guidelines: racism/prejudice (7.1), historical experience and cultural patterns (7.2), positive and negative aspects of ethnic group

experiences (7.3), ethnic diversity (7.5), and the changing nature of ethnic group experiences (7.6). While some may feel that the guidelines listed above are too "political" to be appropriate for large-scale testing programs, we feel that they could provide potential content for item stimuli. Specific objectives 7.4 (1976) may indeed prove to be too sensitive to be included as part of a large-scale instrument (ethnics as active participants in society and as subjects of oppression and exploitation), but its intent can be incorporated into test items which reflect the other specific guidelines. In our judgment, substantive ethnic content can and should be incorporated into the item development/selection processes used by test publishers. The 1971 and 1976 Guidelines should provide an excellent resource for test publishers in selecting item content.

Happily, ethnic tokenism has generally been avoided. But tokenism toward social issues, other than those of an ethnic nature, appears to be the rule rather than the exception. The tokenism of which we speak is both tokenism by omission and tokenism by Commission. Many significant social issues are judiciously avoided (income distribution, social welfare, unemployment, etc.), and others dealt with in a cursory manner (prejudice). We recognize that on a test of limited length a large number of social issues cannot be included on any single test. We do feel that these issues are at least as important (of as much worth?) as clipping fingernails, playing the piano, and the Italian Renaissance.

The development of intellectual skills (what we have called Scope and Sequence-Skills) has traditionally been, and still remains, one of the most important self-identified outcomes of social education. The

difficulties in identifying what these skills are has already been addressed in this report. Test publishers, however, have attempted to identify these skills and include them in a systematic fashion as one of the characteristics of items appearing on their tests. What is patently unclear is why the publishers have even paid attention to intellectual skills. In reporting test results back to users scores are usually related in terms of percentile-ranks for NRT's and master/nonmastery scores for CRT's. We know of no test publisher who reports children's scores on an item by item basis which is representative of the intellectual skills which are assessed on a test.

More importantly, no publisher in any way defines or describes a cognitive skill "roadmap" which may have been used to construct/select items for inclusion on a test. This information could be extremely useful to test users who are interested in obtaining as much information as possible about a child's performance. It may be useful to know not only that a child is at the 43rd percentile in social studies but that she/he answered all of the recall items correctly but only a small proportion of the items at the higher cognitive levels. This presupposes, however, that social studies educators, at whatever level, are familiar with and concerned about these types of outcomes. There is, unfortunately, little evidence that social studies educators are either knowledgeable or concerned beyond a lip-service level. Thus, the size of the market for such a reporting system may not be large enough to justify a publisher's attention to it. Essentially, in the cognitive skills area, we suggest that test publishers consider:

1. Describing and justifying the model or models (if any) of cognitive functioning used to construct test items,

2. Identifying which test items are associated with the components of the model(s), and in what sequence,

3. Reporting results back to test users as to student performance (individual or group) by the cognitive skill assessed,

4. Assisting local school personnel in the interpretation of such scores and in incorporating these results into the instructional program.

Another skill area which has received little attention from test publishers/developers is the ability to distinguish fact from opinion (11.1, 1976). Guidelines dealing with Scope and Sequence-Content could provide a wealth of issues, points of view, and stimulus material for assessing not only 11.1 but many of the skills, explicit or implied which are contained in the Guidelines.

We have included the process of valuing as a part of our factor Scope and Sequence-Skills. Very little evidence can be found in any of the social studies tests which were reviewed that would indicate the affective area receives a high priority in large-scale assessment programs, at least as far as student outcomes are concerned. While many test items are blatantly value-laden and myth promoting, e.g. there is little governmental control in our economic system; there is very little in the way of judging the worth of values or student utilization of a valuing process. One reason may be, of course, that identifying or defending a "correct" answer may be impossible. The only minor exceptions

to this general rule may be found in a few state tests which have "add on" sections that include items of the attitude-survey type. There is apparently no systematic, substantive attempt on these tests to deal with value issues/processes outlined in the Guidelines.

For the most part, test publishers should be complimented for the manner in which they have sought to broadly sample the knowledges representative of social studies and to the extent they seek to measure various cognitive skills. In fact, the COT observed that the better test publishers seemed to be ahead of the profession in the definition of social studies content and skills. But, as pointed out earlier, large-scale assessment has generally been focused upon norm-referenced tests. The graduation competency movement in the various states and the accompanying criterion-referenced (so-called) movement is producing new pressures on social studies educators. The trend of the competency movement is to test for the recall of low level knowledge; in fact knowledge that may or may not be consistent with NCSS Guidelines. Frequently, the stated purposes of these instruments are to test for survival skills and/or basic knowledges. Purported to be criterion-referenced, these tests create the dangerous situation in which students are recycled into classes where "basic" learners are studying under curriculum objectives that would be repugnant to social studies educators. Thus, we may have curriculum by "deflection."

There is another dangerous outcome of the competency testing movement, the loss of curriculum space. Frequently, social studies educators have fought against the creation of competency testing only to discover that students who failed tests in language arts and mathematics were then recycled into remedial classes at the expense of the social

studies classes. There is a simple alternative for the profession; an improved rationale for the social studies, clarity in defining social studies skills and a manageable sequencing of social studies concepts and skills.

The COT realized that the NCSS Guidelines were never intended as descriptions of intended learning outcomes that could be evaluated through commercial tests. Nevertheless, the NCSS Guidelines documents did assert that the guideline statements could be used to evaluate the soundness of social studies programs in the schools, which implied that student behaviors would be impacted. The COT has determined that the Guidelines fall short of an adequate description of social studies behaviors for both the curriculum developer and the test publisher. The question must be asked, "Is the NCSS leadership and membership seriously concerned about evaluation in the social studies?" Do we want to clear up our language in social studies in order that we can identify skill and content domains to be evaluated so that we can give help to curriculum leaders preparing performance outcomes? Certainly, the Guidelines would suggest that the NCSS should be concerned. Guidelines grouped under 4.0 NCSS (1971) Guidelines clearly assert that classroom instruction should be based upon clearly formulated objectives that include knowledge, abilities, valuing and social participation; while 8.0 NCSS (1971) Guidelines go further to assert that evaluation must be based upon these curriculum objectives (intended learning outcomes). Clearly and decisively, the Guidelines augur for the classic curriculum model--carefully worded descriptions (definitions) of the social studies, performance objectives, instructional plans and a consistent plan of student evaluation.

During the development of this report, the COT was at times encouraged and at other times discouraged. We feel that we accomplished (given the conditions) our task in a positive fashion. We now feel that the NCSS must decide to make a serious and concerted effort to improve and influence social studies evaluation. We feel that there are distinct implications for continuation of the current "salutary neglect" on testing. We think that the current trend toward graduation competencies in social studies may "deflect" the social studies curriculum in directions that are professionally undesirable. Without clearly stated NCSS goals, these competency instruments may be developed totally independent of the profession.

Should the NCSS decide to take a more proactive stance on evaluation (which we endorse) we think that there are many positive implications. First, the NCSS could take action to make evaluation processes a major focus of its publications for both the in-service and pre-service social studies teacher. Evaluation could be promoted in Social Education as an integral part of curriculum/instruction planning and not simply the appendage that it often becomes. We suggest that the NCSS should conduct a conference with test publishers to share mutual concerns and to determine appropriate roles for test makers, test users and curriculum developers. Furthermore, we feel that the NCSS should help to bring about more enlightened usage of evaluation processes and data—e.g. the appropriate uses of norm-referenced tests, criterion-referenced tests; diagnostic/prescriptive teaching and the use of test results and the construction and use of classroom tests. Perhaps NCSS could establish a standing committee to advise the states and localities, as well as commercial

publishers, on the development of evaluation instruments. Such a committee could be charged with the responsibility to publish current trends and changes in evaluation procedures in social studies education. NCSS must take an active role in bringing needed information and skills to the profession so we can be better able to accurately evaluate those things we have identified as important outcomes of social education.